UNIVERSIDADE FEDERAL DE GOIÁS (UFG)
INSTITUTO DE FÍSICA (IF)
PROGRAMA DE PÓS GRADUAÇÃO EM FÍSICA (PPGF)

YURI DE JESUS ALVIM

Landauer principle in the context of relativistic communication theory

GOIÂNIA
2023

UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE FÍSICA

TERMO DE CIÊNCIA E DE AUTORIZAÇÃO (TECA) PARA DISPONIBILIZAR VERSÕES ELETRÔNICAS DE TESES

E DISSERTAÇÕES NA BIBLIOTECA DIGITAL DA UFG

Na qualidade de titular dos direitos de autor, autorizo a Universidade Federal de Goiás (UFG) a disponibilizar, gratuitamente, por meio da Biblioteca Digital de Teses e Dissertações (BDTD/UFG), regulamentada pela Resolução CEPEC nº 832/2007, sem ressarcimento dos direitos autorais, de acordo com a Lei 9.610/98, o documento conforme permissões assinaladas abaixo, para fins de leitura, impressão e/ou download, a título de divulgação da produção científica brasileira, a partir desta data.

O conteúdo das Teses e Dissertações disponibilizado na BDTD/UFG é de responsabilidade exclusiva do autor. Ao encaminhar o produto final, o autor(a) e o(a) orientador(a) firmam o compromisso de que o trabalho não contém nenhuma violação de quaisquer direitos autorais ou outro direito de terceiros.

1. Identificação do material bibliográfico

[X] Dissertação   [ ] Tese   [ ] Outro*:_____________

*No caso de mestrado/doutorado profissional, indique o formato do Trabalho de Conclusão de Curso, permitido no documento de área, correspondente ao programa de pós-graduação, orientado pela legislação vigente da CAPES.

Exemplos: Estudo de caso ou Revisão sistemática ou outros formatos.

2. Nome completo do autor

Yuri de Jesus Alvim

3. Título do trabalho

Landauer principle in the context of relativistic communication theory

4. Informações de acesso ao documento (este campo deve ser preenchido pelo orientador)

Concorda com a liberação total do documento [X] SIM   [ ] NÃO[1]

[1] Neste caso o documento será embargado por até um ano a partir da data de defesa. Após esse período, a possível disponibilização ocorrerá apenas mediante:
a) consulta ao(à) autor(a) e ao(à) orientador(a);
b) novo Termo de Ciência e de Autorização (TECA) assinado e inserido no arquivo da tese ou dissertação.
O documento não será disponibilizado durante o período de embargo.
Casos de embargo:
- Solicitação de registro de patente;
- Submissão de artigo em revista científica;
- Publicação como capítulo de livro;
- Publicação da dissertação/tese em livro.

Obs. Este termo deverá ser assinado no SEI pelo orientador e pelo autor.

Documento assinado eletronicamente por **Lucas Chibebe Celeri**, **Professor do Magistério Superior**, em 03/05/2023, às 16:12, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Yuri De Jesus Alvim**, **Discente**, em 03/05/2023, às 16:54, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

A autenticidade deste documento pode ser conferida no site https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **3718974** e o código CRC **BD6F665F**.

Referência: Processo nº 23070.020206/2023-22 SEI nº 3718974

YURI DE JESUS ALVIM

Landauer principle in the context of relativistic communication theory

Dissertação apresentada ao Programa de Pós-Graduação em Física (PPGF), do Instituto de Física (IF), da Universidade Federal de Goiás (UFG), como requisito para obtenção do título de Mestre em Física.

Área de concentração: Física
Linha de pesquisa: Óptica Quântica e Informação Quântica

Orientador: Prof. Dr. Lucas Chibebe Céleri

GOIÂNIA
2023

Ficha de identificação da obra elaborada pelo autor, através do
Programa de Geração Automática do Sistema de Bibliotecas da UFG.

Alvim, Yuri de Jesus
Landauer principle in the context of relativistic communication theory [manuscrito] / Yuri de Jesus Alvim. - 2023.
iv, 109 f.: il.

Orientador: Prof. Dr. Lucas Chibebe Céleri.
Dissertação (Mestrado) - Universidade Federal de Goiás, Instituto de Física (IF), Programa de Pós-Graduação em Física, Goiânia, 2023.
Bibliografia. Apêndice.
Inclui siglas, abreviaturas, símbolos, gráfico, lista de figuras.

1. Information Theory. 2. General Relativity. 3. Quantum Field Theory. 4. Landauer principle. 5. Thermodynamics. I. Céleri, Lucas Chibebe, orient. II. Título.

CDU 53

UNIVERSIDADE FEDERAL DE GOIÁS

INSTITUTO DE FÍSICA

ATA DE DEFESA DE DISSERTAÇÃO

Ata nº 212 da sessão de Defesa de Dissertação de Yuri de Jesus Alvim, que confere o título de Mestre em Física, na área de concentração em Física.

Aos 25 dias do mês de abril de 2023, a partir das 14h00min, por meio de videoconferência, realizou-se a sessão pública de Defesa de Dissertação intitulada “Landauer principle in the context of relativistic communication theory”. Os trabalhos foram instalados pelo Orientador, Professor Doutor Lucas Chibebe Céleri (IF/UFG), com a participação dos demais membros da Banca Examinadora: Professor Doutor Daniel Augusto Turolla Vanzella (IFSC/USP), membro titular externo; e Professor Doutor Rômulo Cesar Rougemont Pereira (IF/UFG), membro titular interno. Durante a arguição, os membros da banca não fizeram sugestão de alteração do título do trabalho. A Banca Examinadora reuniu-se em sessão secreta a fim de concluir o julgamento da Dissertação, tendo sido o candidato aprovado pelos seus membros. Proclamados os resultados pelo Professor Doutor Lucas Chibebe Céleri, Presidente da Banca Examinadora, foram encerrados os trabalhos e, para constar, lavrou-se a presente ata que é assinada pelos membros da Banca Examinadora, aos 25 dias do mês de abril de 2023.

TÍTULO SUGERIDO PELA BANCA

Documento assinado eletronicamente por **Lucas Chibebe Celeri**, **Professor do Magistério Superior**, em 25/04/2023, às 16:08, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Romulo Cesar Rougemont Pereira**, **Professor do Magistério Superior**, em 25/04/2023, às 16:09, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Daniel Augusto Turolla Vanzela**, **Usuário Externo**, em 25/04/2023, às 16:10, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Referência: Processo nº 23070.020206/2023-22 SEI nº 3666294

I dedicate this research to my grandmother, whose presence and teachings were essential for my life.

Acknowledgments

I would like to express my deeply gratitude to my professor and advisor Prof. Dr. Lucas Chibebe Céleri for his invaluable patience and help. Additionally, this work would not have been possible without the generous support from CAPES, who financed my research.

I also would like to mention my family and friends, in specially my cousins and my girlfriend. Their belief in me helped me get through very difficult times during this process.

Abstract

This monograph aims to investigate Landauer's principle for a quantum system in a relativistic context. Specifically, we consider a communication channel described by a mode of a quantum field in a curved spacetime. In order to accomplish that, an introduction to Shannon's theorems, as well as some very important concepts in information theory, will be first presented. The relativistic spacetime structure will be described in the sequence, such that the quantum field theory can be used, in this context, to describe the communication channel. Lastly, the calculation of the channel capacity and the energy contributions for the transmission of information over such a channel will be done. Such calculations provide three different contributions to the total energy variation of the system: one due to the change in the spacetime, one associated with the work necessary to switch on or off the interaction between the detectors —qubits employed to read and codify the information— with the field and, finally, the last one which is due to the communication process itself. This third contribution vanishes for the considered communication channel, so that no extra energy cost is needed to transmit information once the states of the qubits are settled. The original contribution of this work enters here by considering the cyclical conversion of the transmitted information into work. Since no energy is transferred from the sender to the receiver, this engine apparently violates the second law of thermodynamics. By employing Landauer principle we find the energy contribution which allows the receiver to convert information into work without contradictions with thermodynamics.

Keywords: Information Theory; General Relativity; Quantum Field Theory; Landauer Principle; Thermodynamics.

Resumo

Esta monografia tem como objetivo investigar o princípio de Landauer para um sistema quântico em um contexto relativístico. Especificamente, consideramos um canal de comunicação descrito por um modo de um campo quântico em um espaço-tempo curvo. Para isso, uma introdução aos teoremas de Shannon, bem como alguns conceitos muito importantes em teoria da informação, serão apresentados. A estrutura espaço-temporal relativística será descrita na sequência, de modo que a teoria do campo quântico possa ser usada, nesse contexto, para descrever o canal de comunicação. Por fim, o cálculo da capacidade do canal e das contribuições de energia para a transmissão de informação sobre tal canal será realizado. Tais cálculos fornecem três contribuições diferentes para a variação total de energia do sistema: uma devida à mudança na métrica do espaço-tempo, outra associada ao trabalho necessário para ligar ou desligar a interação entre os detectores — qubits usados para ler e codificar a informação — com o campo e, finalmente, a última que é devida ao processo de comunicação em si. Essa terceira contribuição desaparece para o canal de comunicação considerado, de modo que nenhum custo de energia extra é necessário para transmitir informação uma vez que os estados dos qubits são definidos. A contribuição original deste trabalho entra ao considerar a conversão cíclica da informação transmitida em trabalho. Como nenhuma energia é transferida do remetente para o destinatário, este motor aparentemente viola a segunda lei da termodinâmica. Ao empregar o princípio de Landauer, encontramos a contribuição de energia que permite ao receptor converter informação em trabalho sem contradições com a termodinâmica.

Palavras-chave: Teoria da Informação; Relatividade Geral; Teoria Quântica de Campos; Princípio de Landauer; Termodinâmica.

List of Figures

Contents

Chapter 1

Introduction

The main goal of this dissertation is to apply Landauer principle to a communication system in globally hyperbolic spacetimes. Here we will consider a system composed by two qubits (the detectors) and a quantum scalar field, as the communication channel, in a general curved spacetime.

The idea is to mathematically describe the communication channel and calculate its capacity, as well as the energy cost to convey information, so we can apply the Landauer principle to this system and investigate the limits of work extraction on the receiver's laboratory. In order to properly understand the problem at hand, we need to introduce the main mathematical and physical concepts that will be involved.

The first theory we will present here is information theory. This one will be essential so we can understand how a communication process works as well as some very important concepts which will be worked out in this dissertation, as communication over a channel and the channel capacity.

The advancing of computing has enabled a great development of the communication systems, which requires more in-depth scientific expansion. The processes which are associated with data, such as storing, transmission or compression, as well as many others, are investigated by a branch of science we call Information Theory [1].

The theory is committed to investigate the mathematical laws which describes data processing. The interest in information grew up with the creation of electrical telegraphs, which were the first electrical telecommunication system. But the big step towards the modernization of information theory was made with the publication, in 1948, of Claude Shannon's "A Mathematical Theory of

Communication" [1]. He developed the information theory without the semantic aspect of communication, treating codes in an isolated way from the message, such that the actual message would be the one selected from a set of possible ones. In other words, Shannon associated the transmission of information according to the probability distribution associated with the occurrence of an event [1–3]. In Chapter 1 we will talk about the requirements of information measure, which were postulated by Shannon, and this discussion will be more clear.

Additionally, according to Shannon, the communication process consists of five steps. The first one is the production of the message which is intended to be communicated, this step is made by an *information source*. After that we have the process of *encoding*, which is the codification of information into a suitable system so it can be transmitted by the channel. Next, we have to choose the *channel of communication*, which is the medium used to transmit the information from the sender to the receiver, which can be a piece of paper or even a massless field. The fourth step is the *decoding*, that is the reconstruction of the message from the output of the channel. The fifth and last step is executed by a *receiver*, who has the task of reading the decoded message.

Shannon's theory is based on two theorems, the *data compression* and *channel capacity*. The unit for measure information is called *bit* and the encoder uses a certain amount of bits to represent the code which describes the message. That being said, the process of data compression consists of encoding the same message into fewer bits, but in a reliable way. Shannon's data compression theorem determines a fundamental limit to data compression, such that no information is lost in the asymptotic limit [1]. This fundamental rate is known as Shannon entropy.

On the last step of the communication process, it can happen that the receiver reads a message which is different from the one that was intended to. It happens when the communication channel is noisy, i.e., there are some kind of interference which can corrupt the code, thus modifying the message itself. When we consider this kind of channel, there must be a maximum amount of information which can be reliably transmitted. This quantity is called the channel capacity. Shannon's second theorem determines an upper bound for the information which can be reliably conveyed by any noisy channel.

These theorems, which will be presented in the next chapter, are the foundations of what is known today as information theory. By considering that the quantum measurement process is a communication channel, we can extend the classical theory to the quantum domain. In this context, von Neumann entropy takes the place of Shannon's entropy.

There are many applications for information theory, which includes the analysis of molecular sequences and characterizations of resistance mutations [4], channel coding correction, data compression (RAR/ZIP files) and quantum computing [3, 5]. Additionally, the information theory could be applied in problems which includes communication with satellites, space probes and also planes, but it will requires some relativistic analysis, since gravity will have some influence on the results [6, 7]. Thus, depending on the case, we must consider the consequences of the relative motion between the parts which will communicate with each other or even the presence of a black hole [8–11].

Also, since our system is immersed on a globally hyperbolic spacetime, it will be necessary to introduce some ideas about relativity [8, 9] so we can properly define the conditions and limits for our system. To define what are globally hyperbolic spacetimes first we need to understand how to describe the causal structure of spacetime. If we want to investigate the problem of two observers communicating in a curved spacetime we need to know if they can actually communicate with each other. This is why the spacetime causal structure will be important when we explore the communication problem.

However, the causal structure of spacetime depends on concepts like time and space, but these quantities have been defined differently according to the advancement of science. In Newton's conception, the universe is a four-dimensional affine space $\mathbb{A}^4$, where each element of $\mathbb{A}^4$ is called event, and time is defined as a linear map which takes us from the four-dimensional affine space to a one-dimensional one, the real line. We can use the kernel of this map to define what we call simultaneous hypersurfaces, i.e., hypersurfaces where all events on it are simultaneous to each other, and the causal structure of spacetime will be described accordingly with it. But, as we will see, this concept was completely changed when Einstein introduced his relativity theory. In Chapter 3 we will provide a brief idea of Newtonian spacetime and explain why Einstein's theory of general relativity changed all of it.

Additionally, communication implies that the information, that is intended to be sent, needs a medium to be conveyed. In our case, it will be a quantum scalar field. Therefore, the last theory we will need to introduce is the quantum field theory, since it will allow us to describe mathematically the communication channel and many other things which are associated with it, as the channel capacity and the energy cost to convey information over it.

We will consider some concepts of quantum mechanics as well as the spacetime structure described on the first half of Chapter 3 to develop a quantum field theory for the Klein-Gordon scalar field. The field is of great importance to us, since it is used as the communication channel, and we can describe the dynamics of the system by the field equation. At this point we will need to consider globally hyperbolic spacetimes so we can formulate a well posed initial value theory and actually describe the dynamics of the system. Once this process is done, we will be prepared to understand the problem at hand.

In this monograph, we are aiming to apply Landauer principle in the context of relativistic communication. In order to do that, we will first review the calculation of the maximum amount of reliably information that one can transmit through this noisy channel, i.e., the channel capacity, as well as the energy contributions for the transmission of information [6, 12]. Then, we will see that this energy can be separated into three parts, one due to the change of the spacetime metric, other due to the work necessary to switch on an off the qubit, i.e., the basic unit of quantum information, described by a two-level quantum system interacting with the field. This is the last part of the process of communication. But, as we will see on Chapter 4, the last contribution will vanish for states which maximizes the signalling amplitude, such that one could send information without any extra energy cost [11, 13]. To understand the balance of energy and the possibility of converting information into work without violating the second law of thermodynamics is the goal of the present work. In order to do this, we investigate Landauer principle at the receiver's laboratory.

Landauer principle is an important concept in communication theory because it sets a fundamental limit on the amount of energy required to perform a computation or erase a bit of information [14]. This principle has important implications for the design of communication systems and information processing devices.

The principle states that erasing a bit of information necessarily involves dissipating a minimum amount of energy, which is proportional to the product of the temperature and the entropy change associated with the erasure process. This means that there is a fundamental limit to the energy efficiency of any computation or communication process. We will investigate the problem of using the energy dissipated in this process to produce work through a thermal machine and understand what constraint needs to be placed on the system so that it satisfies the second law of thermodynamics [15–19].

In order to understand these calculations, as well as the physical meaning of it, we will introduce some very important concepts. The text will be organized as follows. In Chapter 2 we will present the Shannon's information theory, presenting the different kinds of entropy, as well as the relation between some of them. Also, we will demonstrate Shannon's theorems, which rests in the basis of information theory. In Chapter 3, we will give an introduction about some concepts of general relativity and, in this context, we will develop a quantum field theory. Lastly, in Chapter 4, we will calculate the classical channel capacity, as well as the energy cost of the communication process.

Chapter 2

Shannon information theory

In this chapter we will discuss Shannon's main contributions to information theory, which are *data compression* and *channel capacity* theorems. They were of great importance for the development of some major ideas in communication theory, quantum computation and many other fields. Also, it will be very important so we can understand the next chapters.

Aiming to investigate the theorems proposed by Shannon, we need to comprehend some details about the process of communication. In order to do that, suppose that two scientists, Alice and Bob, want to communicate with each other. Alice, who will send the information, chooses a symbol, say α, she wants to transmit to Bob. After that, this message is *encoded* by an *encoder* $\mathcal{E}$, which is responsible to write the symbol, or any sequence of them, into a specialized format called a *codeword.* The encoding is the process of converting what is intended to be communicated into codes, whose format will depend on the communication channel, e.g., if the communication channel is a letter then the code must be words which can convey the message.

Bob, who is the receiver, will *decode* the codeword in order to read the message. To do that, he uses a *decoder*, $\mathcal{D}$, which implements the opposite process performed by the encoder. It will turn the received code into the message that was intended to be communicated. If the decoded message is the same of that one sent by Alice or not, will depend if the channel, i.e., the medium in which all inputted information was transmitted, is noiseless (ideal) or not.

In information theory, also in computation, the message is encoded in a certain amount of bits. The bit is the smallest unit of information that can be stored and processed, and has two possible values, 0 or 1. For example, if the codeword associated with the symbol α is 00, then we need two bits of information to represent it. But in a real communication Alice wants to send a lot of symbols and each one of them will require a certain amount of bits to be represented.

The process of data compressing is the operation of encoding the message by using as few bits as possible. It can be seen mostly when we convert some archive in the computer to another format, such that the second one occupies less memory, e.g., one can compress a book which is in PDF to ZIP, but, as we will see later, some of the information can be lost during the process.

Shannon's idea, for the noiseless coding theorem, was to achieve the minimal amount of bits that one can encode the output of the information source, i.e., the one from which the information is obtained. Attempting to do that, his proposal was to let the information source emits a large number of symbols and characterizes it as a sequence, which will convey the message. As will be clear soon, the assumption of large numbers will be very important, since we will use it to define an asymptotic limit where the compression rate approaches Shannon entropy. Thus, Shannon also provided an operational interpretation of entropy by proving that it is maximum achievable rate for data compression.

Now, if the channel used by Alice and Bob is noisy, i.e., it can corrupt the message, the information Bob receives is not completely reliable. Therefore, they will need to develop a coding scheme such that Bob receives the correct message. If Bob declares an error every time he receives a wrong message, he could calculate a probability of error. Thus, they could choose a code such that this probability is very small. But, as we will see later, it requires Alice to use much more the communication channel in order to encode the same message into bigger sequences. In contrast with that, this process will decrease more and more the efficiency of the communication, since it is inversely proportional to the number of uses of the channel.

The problem here is that the communication between Alice and Bob will not be successful if the efficiency of the channel is too low. To solve it, Shannon suggested the use of a kind of sequence which takes all the emission probability with several uses of the channel. These sequences are called *typical sequences*. We will talk more about it later in the chapter.

By considering that, Bob knows what kind of sequences will have the majority of the probability of emission, so he could declare an error every time the received sequence is not typical. Therefore, with several uses of the noisy channel, the probability of error can be made arbitrarily small, since the emission probability of typical sequences will approach one. In this way, there must be a maximum amount of information which can be transmitted through the channel reliably, per channel use. Shannon called this upper limit the *channel capacity*. We will see more about it when we discuss his second theorem. But before we go into that, let us first introduce some concepts about the information theory, which will be very important for understanding Shannon's theorems.

2.1 Elements of Information theory

The first concept we aim to introduce is the information measure. In other words, how can we quantify the amount of information in a given source. Here enters the Shannon entropy. Also, we will discuss some other measures, deeply related with the entropy, that will be very useful to prove some of the results in the next chapters. These are the relative entropy, the joint entropy, the conditional entropy and also a measure of correlations, the mutual information. The meaning of all of these quantities will become clear during the development of this chapter.

Before we present Shannon entropy, we need to understand the requirements that any information measure must satisfy, these are known as Shannon's postulates of information theory. The three postulates are [1, 2]:

i. The amount of information in an event, $i(p)$, must depends only upon its probability.

ii. The function which describes the amount of information, $i(p)$ where p is the probability associated with the occurrence of some event, is continuous.

iii. For two independent events, say x and y, with probabilities p_x and p_y, respectively, the information carried in both events together, $i(p_x, p_y)$, is the sum of the information carried by each one of the events, $i(p_x) + i(p_y)$, i.e., it satisfies the additivity property.

These postulates are very important so we can treat information accordingly with probabilities distributions. In this way, the communication scheme operates independently of the outcome, since we do not know which one it will be. The first postulate can be understood by thinking about $i(p)$ as how surprise we are by the occurrence of an event. For example, a tiger hunting in the sea carries much more information than a lion hunting in a Savannah. There is only one kind of function which satisfies all of these requirements, that is the logarithm function. Thus, it must be used when we pretend to measure information.

2.1.1 Shannon Entropy

In a few words, Shannon entropy measures the average amount of information Bob receives from Alice. To define it, let us consider the set χ, of size $|\chi|$, from which an information source can select a symbol x, according to some probability distribution $p_X(x)$. X is the random variable associated with x, i.e., $X = \{x \in \chi, p_X(x)\}$. If Bob receive Alice's signal and confirms that it is actually the correct one, then we say that an event x has occurred.

Therefore, Bob's *surprise* when he read the message can be defined as

$$i(x) \equiv \log\left(\frac{1}{p_X(x)}\right), \tag{2.1}$$

such that, it gets bigger as the probability, $p_X(x)$, of the occurrence of the event x decreases. Thus, we can define the amount of information he gets from reading the symbol x as[2, 3, 5]

$$i(x) \equiv -\log(p_X(x)). \tag{2.2}$$

where the logarithm is taken to base two, since information is measured in bits. This is a very intuitive notion of information. If we want a general theory, any measure of information should depend only on the probabilities of occurrence of the symbols in the set χ. Moreover, we should also expect that the information we get from the occurrence of a given event should increase as its probability of occurrence decreases. Also, it has an interesting, and desired property called *additivity*. Suppose the source selects, independently, two symbols, x_1 and x_2, which are associated, respectively, with the random variables X_1 and X_2. This will occur according to the joint probability $p_{X_1 X_2}(x_1, x_2)$.

The information content of these symbols will be [3]

$$\begin{aligned} i(x_1, x_2) &= -\log(p_{X_1X_2}(x_1, x_2)) && (2.3)\\ &= -\log(p_{X_1}(x_1)p_{X_2}(x_2)) && (2.4)\\ &= -\log(p_{X_1}(x_1)) - \log(p_{X_2}(x_2)) && (2.5)\\ &= i(x_1) + i(x_2). && (2.6)\end{aligned}$$

Actually, by imposing additivity and the other properties discussed above, it is possible to prove that Eq. (2.2) is the unique measure of information.

Now, if we take the average value of the information content of the symbol x, we obtain

$$H(X) \equiv \sum_{x\in\chi} p_X(x)i(x) = -\sum_{x\in\chi} p_X(x)\log(p_X(x)), \tag{2.7}$$

which is defined as the *Shannon entropy*. It can be interpreted as the average information Bob gets once he measures the value of the random variable X that characterizes the information source. Alternatively, we can understand this quantity as the amount of uncertainty we have about a random variable before knowing its value. The first coding theorem provides an operational interpretation for the entropy.

Since the information content is additive, this property will also be satisfied by Shannon entropy. This can be seen by taking the average value of Eq. (2.6).

$$\begin{aligned} \sum_{x_1,x_2\in\chi} p_{X_1X_2}(x_1,x_2)i(x_1,x_2) &= \sum_{x_1\in\chi} p_{X_1}(x_1)i(x_1) + \sum_{x_2\in\chi} p_{X_2}(x_2)i(x_2),\\ H(X_1,X_2) &= H(X_1) + H(X_2). && (2.8)\end{aligned}$$

Another property satisfied by Shannon entropy is that it is maximized when the probabilities $p_X(x)$ are uniform. To prove it, consider the uniform distribution

$$p_X(x) = \frac{1}{|\chi|}. \tag{2.9}$$

The Shannon entropy will be

$$
\begin{aligned}
H(X) &= -\sum_{x\in\chi}^{|\chi|} p_X(x)\log(p_X(x)) \\
&= -\sum_{x\in\chi}^{|\chi|} \frac{1}{|\chi|}\log\left(\frac{1}{|\chi|}\right) \\
&= -|\chi|\frac{1}{|\chi|}\log\left(\frac{1}{|\chi|}\right) \\
&= \log(|\chi|). \qquad (2.10)
\end{aligned}
$$

Since the minimum amount of information which Bob can get, by determining the arrived signals, is zero, i.e., $i(x) \geq 0$, the Shannon entropy will be limited between 0 and the $\log$ of the dimension of the system, which, in this case, is $|\chi|$. This property is called *positivity.*

As an example, suppose we want to calculate the Shannon entropy of the result of tossing a coin. If each side of it has the same probability,

$$p_X(x) = \frac{1}{2}, \qquad (2.11)$$

of coming up, Shannon's entropy will be

$$H(X) = -2\frac{1}{2}\log\left(\frac{1}{2}\right) = 1. \qquad (2.12)$$

Now, considering the case where the probabilities are not uniform, e.g., the probability of getting a tail is 0.2, then the entropy becomes

$$H(X) = -0.2\log(0.2) - 0.8\log(0.8) = 0.72 < 1. \qquad (2.13)$$

Thus, as we said before, when the probabilities are uniform the Shannon entropy will be maximized. This can also be formally proved by extremizing the entropy under the restriction of the conservation of probability.

Now that we have introduced the Shannon entropy and discussed some of its properties, we can use it to define other measures of information. Since these measures will describe the relationship between two random variables, we will consider a second one, Y, associated with the symbol y, which has the probability distribution given by $p_Y(y)$.

The first kind of measure we are going to talk about is the relative entropy. It is a type of statistical distance, which describes how much one probability distribution, say $p_X(x)$, differs from another one, say $p_Y(y)$. Then, we will introduce the joint and conditional entropies, which are, the sum of the Shannon entropies of two independent random variables, X and Y, and the amount of information needed to describe one random variable, Y once we know the outcome of another variable X, respectively. Lastly, we will introduce the mutual information, which gives us how much information Bob can have about Alice given that he learned the outcome of Y, i.e., the amount of information a random variable carries about another [2].

2.1.2 Relative entropy

Given the random variables X and Y, associated with the symbols x and y, selected accordingly with the probability distributions $p_X(x)$ and $p_Y(y)$, respectively, we define the *relative entropy* as [2, 3]

$$H(X||Y) \equiv -\sum_{x,y} p_X(x)\log(p_Y(y)) - H(X). \tag{2.14}$$

Also, by using Eq. (2.7), it can be written as

$$H(X||Y) = \sum_{x,y} p_X(x)\log\left(\frac{p_X(x)}{p_Y(y)}\right). \tag{2.15}$$

As we can see from the above equation, the relative entropy will measure how much the probability distribution $p_Y(y)$ differs from a second one $p_X(x)$. Note that, if the support of the probability distribution over X is not contained in that over Y, this quantity gives an infinite value. For this reason, it is also called a divergence. Additionally, the relative entropy $H(X||Y)$ measures how much extra information is needed to encode samples from Y using a code optimized for X, compared to encoding them using a code optimized for X [2]. If we have a model that estimates the probability distribution of some data, but the model is imperfect and the true distribution is slightly different, we can use relative entropy to measure the difference between the estimated distribution and the true distribution.

Moreover, we can state an interesting property that is satisfied by the relative entropy, which is that $H(X||Y)$ is non-negative, vanishing only if $Y = X$. But as the difference between these two random variables increases, the relative entropy will become greater, always assuming positive values [2]. In order to prove the positivity property of Eq. (2.15), we can rewrite it as

$$\sum_{x,y} p_X(x) \left[\log\left(p_X(x)\right) - \log\left(p_Y(y)\right)\right] = \sum_{x,y} \left(-\log \frac{p_Y(y)}{p_X(x)}\right) p_X(x), \tag{2.16}$$

and, since $p_Y(y)/p_X(x) \geq 0$, the negative of the logarithmic function will be convex. Thus, we can use the Jensen's inequality, which states that, for any convex function f, we have

$$f\left(\sum_i p_i x_i\right) \leq \sum_i p_i f(x_i). \tag{2.17}$$

Therefore,

$$\begin{aligned} H(X||Y) &= \sum_{x,y} \left(-\log \frac{p_Y(y)}{p_X(x)}\right) p_X(x) \geq -\log\left(\sum_{x,y} \frac{p_Y(y)}{p_X(x)} p_X(x)\right) \\ &\geq 0. \end{aligned} \tag{2.18}$$

Which proves the positivity of the relative entropy.

2.1.3 Joint entropy

As the name says, the joint entropy is simply the sum of the expected information of two independent random variables, i.e., the amount of information gained from the occurrence of the two events, x and y. The random variable Y can be associated, for instance, with the selection of a second symbol by the information source. If the information source selects each symbol independently, we can define the joint entropy as [2, 3]

$$H(X, Y) = H(X) + H(Y), \tag{2.19}$$

where $H(X)$ and $H(Y)$ are written according to (2.7).

But when X and Y are not independent, we need to consider the joint probability distribution $p_{X,Y}(x, y)$ to calculate this measure, thus taking into account the correlations between the random variables. Then, the joint entropy will be determined by [2]

$$H(X, Y) \equiv -\sum_{x,y} p_{X,Y}(x, y) \log(p_{X,Y}(x, y)). \tag{2.20}$$

$p_{X,Y}(x,y)$ gives the probability of occurrence of the two events x and y, such that, if we consider $p_X(x)$ as the probability of event x to occur, then we can calculate the joint probability by

$$p_{X,Y}(x,y) = p_X(x)p_{Y,X}(y|x), \tag{2.21}$$

where $p_{X,Y}(y|x)$ is the conditional probability $\Pr(Y = y|X = x)$, i.e., the probability of $Y = y$ given that $X = x$.

Therefore, we can write the joint entropy as

$$\begin{aligned} H(X,Y) &= \sum_{x,y} p_{X,Y}(x,y) \log\left(\frac{1}{p_X(x)p_{Y,X}(y|x)}\right) \\ &= \sum_{x,y} p_{X,Y}(x,y) \log\left(\frac{1}{p_X(x)}\right) + \\ &+ \sum_{x,y} p_{X,Y}(x,y) \log\left(\frac{1}{p_{Y,X}(y|x)}\right) \\ &= -\sum_{x} p_X(x) \log(p_X(x)) - \\ &- \sum_{x} p_X(x) \sum_{y} p_{Y,X}(y|x) \log(p_{Y,X}(y|x)), \end{aligned}$$

which gives us the equation

$$H(X,Y) = H(X) - \sum_{x} p_X(x) \sum_{y} p_{Y,X}(y|x) \log(p_{Y,X}(y|x)). \tag{2.22}$$

As we will see, the second term of the right side of the equation is the conditional entropy.

2.1.4 Conditional entropy

We can define the conditional entropy as the amount of information obtained by measuring the random variable Y if we already know the outcome of a second random variable X. It can be written as [2, 3]

$$H(Y|X) \equiv -\sum_{x} p_X(x) \sum_{y} p_{Y,X}(y|x) \log(p_{Y,X}(y|x)). \tag{2.23}$$

where $p_{Y,X}(y|x)$ is the probability of obtaining $Y = y$, given that we already know that $X = x$. Note that, if X and Y are completely independent, then $p_{Y,X}(y|x) = p_Y(y)$, so the conditional entropy will be equal to $H(Y)$. Also, if X and Y are completely correlated, i.e., the value of Y is fully determined by the value of X, the conditional entropy will be zero, since $p_{Y,X}(y|x) = 1$ and, consequently, $\log(p_{Y,X}(y|x)) = 0$.

We can substitute (2.23) into (2.22), to write

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

which relates the joint entropy $H(X,Y)$ of the random variables X and Y, with the Shannon entropy $H(X)$ and the conditional entropy $H(Y|X)$.

Therefore, the conditional entropy is a measure of the change in the entropy of a random variable when we know the value of a second random variable.

2.1.5 Mutual information

The final information measure we will talk about is the *mutual information.* It can be thought as the amount of information the variable X has about Y, or vice-versa. We define the mutual information of two random variables X and Y as [2, 3]

$$I(X:Y) \equiv H(X) - H(X|Y). \tag{2.25}$$

We can use the results of the previous section in order to rewrite the mutual information as

$$\begin{aligned} I(X:Y) &= H(X) - H(X,Y) + H(Y) \\ &= H(Y) - [H(X,Y) - H(X)] \\ &= H(Y) - H(Y|X) \\ &= I(Y:X), \end{aligned} \tag{2.26}$$

which shows that the mutual information is symmetric.

If X and Y are completely uncorrelated, then the mutual information will be zero, since we will learn nothing from one variable given that we know the other one. Another way of thinking about the mutual information is that it is the amount of information the random variables X and Y hold in common, i.e. the correlations between them.

These measures can be better understood if we look at the illustration in Fig. 2.1, where we consider two sets, which are the Shannon entropies $H(X)$ and $H(Y)$. The sum of these sets will give us a greater one, that is known as the joint entropy $H(X,Y)$.

Note that we have an intersection between $H(X)$ and $H(Y)$, which constitutes the common points to both sets, i.e., the mutual information between them. Lastly, we can establish the conditional entropies by looking at the Eq. (2.25), or Eq. (2.24), which gives us the sets $H(X|Y)$ and $H(Y|X)$.

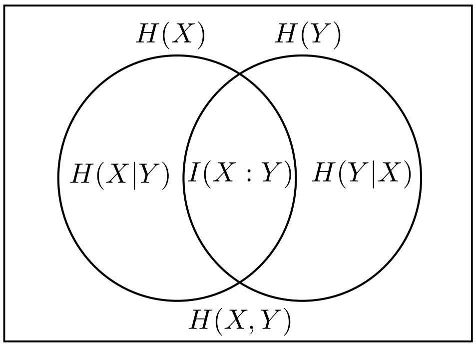

Figure 2.1: Given the random variables X and Y, the relations between the associated information measures are described by the joint entropy, conditional entropy and also the mutual information. From this illustration we can see how these measurements are related to each other. (Source: V. Vedral et al., *Introduction to quantum information science*, 2006)

Now that we presented some basics of information measures, we made our first step in order to understand Shannon theorems. The next concepts we will investigate are data compression and typical sequences. In the next section we give an example of data compression, so that it can be better understood when we present Shannon's noiseless code theorem later. Also, we will discuss about typical sequences, that were cited before, since these ones, together with the law of large numbers, is a very powerful tool to demonstrate the theorems.

2.2 An Example of Data Compression

The first theorem we are going to present here is Shannon's data compression theorem. But before we introduce it properly, we will work on a very interesting example, which will also give us an idea of codification. As before, suppose that Alice and Bob want to communicate with each other by using a noiseless bit channel (a communication channel that takes bits as input and output). The sender, Alice, will need to encode the message she wants to transmit in bits, since the bit channel only accepts bits as input. Then, the channel will carry the information to the receiver, Bob, who will decode and read it. Additionally, since we are considering a noiseless channel, the decoded message will be exactly the same as the one sent by Alice.

For now, suppose Alice chose the symbol "α". It will be encoded into a codeword "00" by the encoder $\mathcal{E}$. Note that the chosen codeword has two bits of information, each one represented by "0". The next step of the communication procedure is the choice of the channel. As illustrated in the Fig. 2.2, they will use a noiseless bit channel, i.e, a communication channel that transmits bits without corrupting them, in order to convey the message.

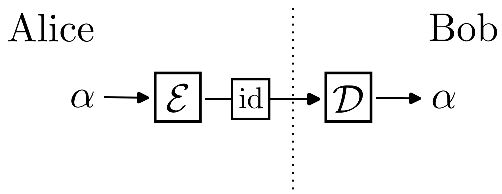

Figure 2.2: Alice encodes the message "α", using the encoder $\mathcal{E}$, to a codeword "00" which is transmitted by the noiseless bit channel (indicated by "id") to Bob. The last step decodes the same message sent by Alice.

Once it is done, Bob will receive the codeword and then decode it, he uses a decoder $\mathcal{D}$. Since we are considering a noiseless channel, for now, he receives the exact codeword "00", which gives the symbol "α" once it is decoded. The communication protocol happens according to Fig. 2.2. It will be similar if Alice chooses more than one symbol.

Therefore, let us consider that Alice wants to send a message by using the set of symbols $\chi = \{\alpha, \beta, \gamma, \delta\}$. An information source can select each symbol according to some probability distribution, which, in this case, will be [3]

$$\begin{aligned} p(\alpha) &= 1/8, \\ p(\beta) &= 1/2, \\ p(\gamma) &= 1/8, \\ p(\delta) &= 1/4. \end{aligned} \tag{2.27}$$

We assume that each symbol is selected independently of the previous ones and after the selection Alice converts it to the appropriate codeword [3]. The message encoding will occur according to a *coding scheme*, i.e., an arrangement of the message as a certain number of bits. In the present case, in which we only have four symbols, she could use the following coding scheme

$$\begin{aligned} \alpha &\rightarrow 00; \\ \beta &\rightarrow 11; \\ \gamma &\rightarrow 01; \\ \delta &\rightarrow 10. \end{aligned} \tag{2.28}$$

Where, the symbol "α" has been encoded into a codeword "00", which has two bits of information, the same with "β", which has been encoded into the codeword "11" and so on. We can verify the performance of this coding scheme by calculating the expected length of a codeword [3]

$$\bar{l} = \sum_{k \in \chi} p(k) l(k), \tag{2.29}$$

with $l(k)$ being the size, in bits, of the codeword used to encode the symbol k. The lower this number is, more efficient will be the coding scheme used by Alice, since the message will be more *compressed* and more information can be conveyed with the same number of channel uses. In the case shown in Eqs. (2.28) it is clear that the expected length of the codeword is two bits, as all symbols are encoded using two bits of information.

Is there a way to devise a more efficient code scheme? The answer, in this case, is yes. To do that, instead of using the same number of bits for each codeword, Alice could encode the symbols by using less bits for the ones with greater probability of occurrence and more bits for symbols less likely to appear, i.e., rare events will carry more information than those that are common. One option is the following code scheme [3]

$$\begin{aligned} \alpha &\rightarrow 110; \\ \beta &\rightarrow 0; \\ \gamma &\rightarrow 111; \\ \delta &\rightarrow 10. \end{aligned} \tag{2.30}$$

The expected length of this code scheme can be determined using Eq. (2.29), so that

$$\bar{l} = 3\frac{1}{8} + 1\frac{1}{2} + 3\frac{1}{8} + 2\frac{1}{4} = \frac{7}{4}, \tag{2.31}$$

which is smaller than the one we found in the scheme (2.28), and consequently, the scheme (2.30) is more efficient.

The process we just presented is called *data compression* or *source coding*. It is the method of encoding information by using fewer bits than before, as we did here. First we had a scheme with an expected length equal to 2, but then we were able to decrease it to 1.75. This process is important because, by using it, we can reduce the amount of resources needed to store or transmit the information.

We can reach another interesting fact by calculating the entropy of the information source. From the probability distribution in Eq. (2.7), we get [3]

$$H = -\frac{1}{8}\log\frac{1}{8} - \frac{1}{2}\log\frac{1}{2} - \frac{1}{8}\log\frac{1}{8} - \frac{1}{4}\log\frac{1}{4} = \frac{7}{4}, \tag{2.32}$$

which is exactly the same as the expected length of the codeword of the above coding scheme. We will see that this is not a coincidence and the above code is optimal, in the sense that this is the maximum achievable rate of data compression.

2.3 Shannon Compression and Typical Sequences

The problem of compressing information is that there is a limit above which, if we keep compressing, we loose some information. Thus, Alice needs to know by how much she can compress the information such that it is possible for Bob to read it. In other words, she needs a maximum achievable rate for compression, so they could communicate with each other reliably.

To solve this problem Shannon proposed one of the fundamental ideas of information theory, which is known as Shannon's compression theorem. Considering a set of symbols χ, with size $|\chi|$, from which the information source will randomly select a symbol, x, associated with a random variable, X, according to the probability $p_X(x)$, i.e., $X = \{x \in \chi, p_X(x)\}$, such that we can write the Shannon entropy as given in Eq. (2.7). Shannon's data compression theorem can be stated as [3]

Theorem 1. *The entropy of an information source, $H(X)$, specified by a discrete random variable X, is the maximum achievable rate for compression.*

In order to demonstrate it, Shannon's idea was to let the source emits a large number of symbols and then characterize it as a block sequence. By doing that, we can codify the whole sequence instead of each symbol separately. Also, this assumption will allow, as we will see later, for us to take the asymptotic limit of the theory, which is essential to demonstrate both theorems.

Therefore, we must consider that the source emits n symbols, selected from the set χ. We define a message as the block sequence [3]

$$x^n \equiv x_1 x_2 ... x_i ... x_n, \tag{2.33}$$

with x_i, $(i = 1, ..., n)$, being the ith emitted symbol. Now here enters the second of Shannon's assumption, that the source emits each symbol in an independently and identically distributed way (i.i.d.), i.e., the selection of the next symbol does not depend on the one selected before and all the symbols are taken under the same probability distribution. We denote by X^n the random variable associated with the messages of length n.

By using the i.i.d. assumption, the probability that the source emits the sequence x^n can be written as [3]

$$\begin{aligned} p_{X^n}(x^n) &= p_{X_1,X_2,...,X_n}(x_1, x_2, ..., x_n) \\ &= p_{X_1}(x_1)p_{X_2}(x_2)...p_{X_n}(x_n) \\ &= p_X(x_1)p_X(x_2)...p_X(x_n) \\ &= \prod_{i=1}^{n} p_X(x_i). \end{aligned} \tag{2.34}$$

The first equality is the general probability of the block sequence. The second one holds because each event is independent of the others, while the third equality comes from the fact that the random variables X_i are identically distributed, so we can choose X to represent all of them.

But making the calculation of the probabilities by using the Eq. (2.34) can be a hard task, since the sequence (2.33) can have a large number of symbols, much larger than the set χ itself. In order to simplify the calculation, we can characterize the symbols by a set of letters a_i, with $i = 1, ..., |\chi|$, such that we can count the number of repetitions of each symbol in the sequence x^n by the quantity $N(a_i|x^n)$. Therefore, the probability given in Eq. (2.34) can be written in terms of the letters a_i and the number $N(a_i|x)$ as [3],

$$p_{X^n}(x^n) = \prod_{i=1}^{n} p_X(x^i) = \prod_{i=1}^{|\chi|} p_X(a_i)^{N(a_i|x^n)}. \tag{2.35}$$

To illustrate this process, suppose that Bob receives and read the sequence $x^n = \beta\beta\delta\gamma\beta\alpha$. In this sequence, the β symbol appears 3 times, so $N(\beta|x^n) = 3$. Also, δ does not repeat, thus $N(\delta|x^n) = 1$ and so on. First, using Eq. (2.34), with the probabilities given in Eq. (2.27), we can compute the probability $p_{X^n}(x^n)$ as

$$p_{X^n}(x^n) = \prod_{i=1}^{n} p_X(x^i) = \left(\frac{1}{2}\right)\left(\frac{1}{2}\right)\left(\frac{1}{4}\right)\left(\frac{1}{8}\right)\left(\frac{1}{2}\right)\left(\frac{1}{8}\right) = \frac{1}{2048}. \tag{2.36}$$

Now, by using Eq. (2.35), we obtain

$$p_{X^n}(x^n) = \prod_{i=1}^{|\chi|} p_X(a_i)^{N(a_i|x^n)} = \left(\frac{1}{2}\right)^3\left(\frac{1}{4}\right)^1\left(\frac{1}{8}\right)^1\left(\frac{1}{8}\right)^1 = \frac{1}{2048}, \tag{2.37}$$

which confirms that we have the same result. However, note that the first method is much less efficient if we have a very large number of symbols.

Moreover, according to statistics, if we have n random variables, as it is the case, we can define the sample mean by [20]

$$\overline{X} \equiv \frac{1}{n}\sum_{i=1}^{n} X_i, \tag{2.38}$$

where $X_1, X_2, ..., X_n$ are i.i.d random variables. Now, note that, if X is a set of random variables of size n, the functions that depend on X will also be random variables. Therefore, replacing X_i with the random variables $-\log(p_X(x_i))$ in Eq. (2.38), the sample mean will be

$$\begin{aligned} -\frac{1}{n}\sum_{i=1}^{n}\log(p_X(x_i)) &= -\frac{1}{n}\log\left(\prod_{i=1}^{n} p_X(x_i)\right) \\ &= -\frac{1}{n}\log(p_{X^n}(x^n)), \end{aligned} \tag{2.39}$$

where we used Eq. (2.34). The above quantity is known as the *sample entropy* [2, 3]

$$\overline{H}(x^n) \equiv -\frac{1}{n}\log(p_{X^n}(x^n)), \tag{2.40}$$

which is the information content of the sequence received by Bob per use of the channel. In other words, if the sample is composed by the block sequence x^n, which is associated with the probability $p_{X^n}(x^n)$, that is also a random variable, the sample mean corresponding to these probabilities is the amount of information Bob will get after learning the sequence x^n per number of symbols, n. This quantity is also known as *compression rate*.

We can rewrite the sample entropy by using Eq. (2.35), as

$$\begin{aligned} -\frac{1}{n}\log(p_{X^n}(x^n)) &= -\frac{1}{n}\log\left[\prod_{i=1}^{|\mathcal{X}|} p_X(a_i)^{N(a_i|x^n)}\right] \\ &= -\frac{1}{n}\sum_{i=n}^{|\mathcal{X}|}\log\left[p_X(a_i)^{N(a_i|x^n)}\right] \\ &= -\sum_{i=1}^{|\mathcal{X}|}\frac{N(a_i|x^n)}{n}\log(p_X(a_i)). \end{aligned} \tag{2.41}$$

It will be clear now why Shannon's assumption is so important. If we let the source emits a large number of symbols, i.e., when n is large enough, according to probability theory, the law of large numbers states that the empirical distribution, $N(a_i|x^n)/n$, of a random sequence x^n approaches to the true distribution $p_X(a_i)$ [3]. This is known as the asymptotic limit. Considering $\Delta > 0$, we can write this statement as [3]

$$\lim_{n\to\infty} p\left\{\left|-\frac{1}{n}\log(p_{X^n}(x^n)) + \sum_{i=1}^{|\mathcal{X}|} p_X(a_i)\log(p_X(a_i))\right| \leq \Delta\right\} = 1, \tag{2.42}$$

but note that the summation in the left side of the equation is the negative of the Shannon entropy $H(X)$, defined in Eq. (2.7), such that

$$\lim_{n\to\infty} p\left\{\left|-\frac{1}{n}\log(p_{X^n}(x^n)) - H(X)\right| \leq \Delta\right\} = 1. \tag{2.43}$$

The above equation says that the compression rate, given by Eq. (2.40), approaches Shannon entropy, given by Eq. (2.7), in the asymptotic limit, which proves the data compression theorem. Note that, without his assumption of letting the source emits a large number of symbols, we could never use the law of large numbers and, consequently, proving the theorem.

Also, based in the asymptotic limit, we can define a kind of sequence that will be fundamental in order to demonstrate the next Shannon's theorem, i.e., the channel capacity. The idea behind the compression scheme is that, in this limit, only an exponentially small number of sequence (compared with the total number of sequences) have all the probability. For this reason, these sequences are called *typical sequences*. In the asymptotic limit, the probability that the source emits one of such sequences approaches one, such that, if we compress only the typical sequences, the error probability, i.e., the probability that Bob receives an atypical sequence, goes to zero as n becomes large. This is the strategy we will use to prove the next theorem, since Alice and Bob will no longer work as a noiseless channel, but as a noisy one.

In order to demonstrate his theorem, Shannon had no other choice than to define typical sequences. Therefore, we can describe a typical sequence and a typical set through the following definitions [3].

Definition 1. *For some $\Delta > 0$, a sequence x^n is said a Δ-typical if its sample entropy $\overline{H}(x^n)$ is Δ-close to the Shannon entropy $H(X)$ of the random variable X.*

Definition 2. *The Δ-typical set $T_{\Delta}^{X^n}$ is the set of all Δ-typical sequences x^n*

$$T_{\Delta}^{X^n} \equiv \{x^n : |\overline{H}(x^n) - H(X)| \leq \Delta\}. \tag{2.44}$$

There are some interesting properties which are satisfied by the typical set $T_{\Delta}^{X^n}$. These are [3]:

1. **Unit probability**. As n becomes large, the sample entropy approaches to the Shannon entropy. Therefore, the probability that an emitted sequence is typical approaches to one.

2. **Smaller Cardinality**. The size of a typical set $|T_{\Delta}^{X^n}|$ is approximately $2^{nH(X)}$ and is exponentially smaller than the size of the set of all sequences $2^{n\log(|\chi|)}$, whenever the random variable X is not uniform.

3. **Equipartition**. The probability of a typical set is approximately uniform, with value $2^{-nH(X)}$.

Since these properties make use of the asymptotic limit, these three together are known as *Asymptotic Equipartition Property* (AEP) [3].

The first property is given in Eq. (2.43). The third one can be proved by manipulating the definition of the typical sequence

$$|\overline{H}(x^n) - H(X)| \leq \Delta. \tag{2.45}$$

Thus, using Eq. (2.40), we have

$$\left| -\frac{1}{n} \log(p_{X^n}(x^n)) - H(X) \right| \leq \Delta, \tag{2.46}$$

which can be rewritten as

$$H(X) - \Delta \leq -\frac{1}{n} \log(p_{X^n}(x^n)) \leq H(X) + \Delta. \tag{2.47}$$

Multiplying the whole equation by $-n$, we obtain

$$-n\left[H(X) + \Delta\right] \leq \log(p_{X^n}(x^n)) \leq -n\left[H(X) - \Delta\right], \tag{2.48}$$

such that,

$$2^{-n[H(X)+\Delta]} \leq p_{X^n}(x^n) \leq 2^{-n[H(X)-\Delta]}. \tag{2.49}$$

Therefore, with the above equation we prove that the probability of a typical set is approximately uniform, i.e., $p_{X^n}(x^n) \approx 2^{-nH(X)}$.

Now, in order to prove the smaller cardinality property, consider the set of all sequences χ^n. The summation of the probabilities of all sequences to be emitted is equal to one, which is obviously greater than the probabilities of the sequences in the typical set. Thus [3]

$$\sum_{x^n \in \chi^n} p_{X^n}(x^n) = 1 \geq \sum_{x^n \in T_\Delta^{x^n}} p_{X^n}(x^n). \tag{2.50}$$

According to the left side of Eq. (2.49), for $\Delta > 0$, the probability $p_{X^n}(x^n)$, associated with typical sequences, is greater than or equal to $2^{-n[H(X)+\Delta]}$. Moreover, since the probability is approximately uniform, the summation will be simply the size of the typical set, $|T_\Delta^{X^n}|$, times the probability of one sequence [3]

$$1 \geq \sum_{x^n \in T_\Delta^{x^n}} p_{X^n}(x^n) \geq \sum_{x^n \in T_\Delta^{X^n}} 2^{-n(H(X)+\Delta)} = |T_\Delta^{x^n}| 2^{-n(H(X)+\Delta)}, \tag{2.51}$$

so, we get

$$|T_{\Delta}^{X^n}| \leq 2^{n(H(X)+\Delta)}. \tag{2.52}$$

Using the right-hand side of Eq. (2.49), and the Unit probability property, we have

$$1 = \sum_{x^n \in T_{\Delta}^{X^n}} p_{X^n}(x^n) \leq \sum_{x^n \in T_{\Delta}^{X^n}} 2^{-n(H(X)-\Delta)} = |T_{\Delta}^{X^n}| 2^{-n(H(X)-\Delta)}, \tag{2.53}$$

and, thus, for sufficiently large n,

$$|T_{\Delta}^{X^n}| \geq 2^{n(H(X)-\Delta)}. \tag{2.54}$$

The size of typical set is smaller than the size of all sequences when the probability distribution is uniform, in which case the entropy $H(X)$ will be equal to $\log|\chi^n|$ by Eq. (2.10). Thus, the size of the typical set will be exponentially smaller than the size of all sequences [3]

$$|T_{\Delta}^{X^n}| \leq 2^{n(H(X)+\Delta)} < 2^{n(\log|\chi^n|+\Delta)}. \tag{2.55}$$

Therefore, we prove the AEP. Now we are ready to understand and demonstrate the *channel capacity theorem*, which is another very important contribution to information theory presented by Shannon.

2.4 Channel Capacity Theorem

In the process of communicating with someone, several problems may arise in order to decrease the effectiveness of communication, such that, the communication is not completely reliable. We experienced a lot of these problems in the pandemic times, where people had to stay in their houses in order to prevent the spread of coronavirus infection (COVID-19), and the communication took place primarily through electronic devices.

Some of these problems could be a bad internet connection, a hardware malfunction, someone trying to intercept the communication and many others. In one way or another they were always there, and anyone can be subject to it, even Alice and Bob. When the communication channel has these problems, we call it a noisy channel. From now on, to investigate Shannon's channel capacity theorem, let us suppose that Alice and Bob will communicate with each other by using a noisy channel.

2.4.1 Noisy Channel

In order to understand what is a channel capacity, first we need to formally introduce what we mean by noisy channel, since it will play a fundamental role in the development of the theory. Thus, consider that Alice and Bob are going to communicate with each other through a noisy channel. In practical terms, it works in the following way. Alice encodes the message "α" into a codeword "0" and sends it to Bob, who will decode it. But, since this is a noisy channel, there's a chance the codeword gets corrupted on its way down the channel, such that Bob receives a different one, which is "1". In such situation, we can calculate the probability of error, which we call p. Therefore, $(1-p)$ is the probability of success of the communication process. The goal is to find a way to maximize this probability.

Because of the probability of error, we can say that the information sent by Alice is not completely reliable. In an attempt to make this probability vanishes, such that the communication could be reliably done, Shannon proposed several uses of the noisy channel, which will allow us to consider the asymptotic limit and, consequently, the typical set unit probability described above, to make the error probability arbitrarily small.

In order to do that, Alice can encode the information into another code, such that Bob can have a higher probability to read the true message. For example, consider that she uses the following scheme [3]

$$0 \to 000; \quad 1 \to 111. \tag{2.56}$$

Therefore, to send the bit "0" or "1", Alice will use the channel three times, such that the rate of communication, i.e., the number of bits bits per channel use decreases by a factor of $1/3$.

If $1-p$ is the probability that Bob receives the true message sent by Alice with a channel use, then $(1-p)^3$ will be the probability of no error at all, i.e., Bob receives the exactly codeword indicated by the code scheme (2.56). The probability of a single-bit error is equal to $3p(1-p)^2$, where the number 3 is because the error can happen for any of the three uses of the channel. For double and triple errors, the probabilities will be, respectively, $3p^2(1-p)$ and p^3 [3].

If Bob admits that the repeated bits are those of the real message, he could be right for the case of no error and a single-bit error, and will be definitely wrong with double or triple-bit errors. So, by using this code scheme there is still a chance of error of

$$p_e = 3p^2(1-p) + p^3 = 3p^2 - 2p^3. \tag{2.57}$$

This method will reduce the error probability if $p_e < p$, as

$$3p^2 - 2p^3 < p, \tag{2.58}$$

then, we can make

$$0 < -3p^2 + 2p^3 + p \tag{2.59}$$

$$0 < p(2p-1)(p-1), \tag{2.60}$$

so, when $0 < p < 1/2$ the probability of error is reduced. Thus, as we can see, for the process to work, there can not be too much noise in the channel, such that the error probability is below $1/2$ [3].

Note that, there is still a chance of error in this scheme, but it can be reduced if Alice associate the bit of information "0" or "1" to a larger codeword, and she could keep doing so until the error probability is negligible, which implies much more uses of the channel. But, as she does it, the efficiency of the channel, i.e., the rate of communication, will also be reduced, since she will use the channel much more times in order to send the same message.

For the communication to be as efficient as possible we need to have a good rate of communication and, according to our previous code, the redundancy scheme takes it to zero as we try to make the error probability arbitrarily small. To solve this problem, Shannon proposed that Alice and Bob should use only typical sequences in the communication process, i.e., Bob declares an error every time he receives a sequence which is not typical. By doing that, one could assume several uses of the channel and make the error probability arbitrarily small. This idea suggests a maximum rate in which the communication process can be done reliably. This quantity is called *channel capacity*.

In order to define a general noisy channel, let us suppose that Alice selects a random message m from a set $\mathcal{M} \equiv \{1, ..., |\mathcal{M}|\}$. Also, as we saw earlier, the maximum Shannon entropy will be equal to $\log$ of the system's dimension. Since the dimension is $|\mathcal{M}|$, the number of bits required to represent the whole set $\mathcal{M}$ is $\log(|\mathcal{M}|)$ [3]. In this way, we can define the rate of communication as

$$R = \frac{1}{n} \log(|\mathcal{M}|), \tag{2.61}$$

where n is the number of channel uses.

The entire communication process can be described as follows. The message m will be encoded into a codeword $x^n(m)$ by the encoder $\mathcal{E}$. Let us call $\mathcal{N}$ the noisy channel, which will corrupt the sequence $x^n(m)$ to $y^n(m)$, delivered to Bob. This process is illustrated in Fig. 2.3. The corrupted sequence y^n is then decoded by the decoder $\mathcal{D}$, producing an estimate message $\hat{m}$, as illustrated in the figure. Moreover, we can associate the sequences $x^n \equiv x_1 x_2 ... x_n$ and $y^n \equiv y_1 y_2 ... y_n$ to the random variables $X^n \equiv X_1 X_2 ... X_n$ and $Y^n \equiv Y_1 Y_2 ... Y_n$, respectively.

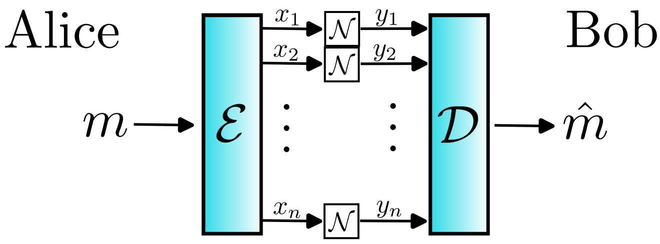

Figure 2.3: Alice uses an encoder $\mathcal{E}$ to encode the message m into the sequence x^n, which will be corrupted by the noisy channel $\mathcal{N}$, generating the corrupted sequence y^n. In the last step, Bob decodes the message throughout the decoder $\mathcal{D}$, obtaining the message $\hat{m}$.

Clearly, the sequence y^n will exist only if Alice encodes her message. In this way, the probability of the output sequence is a conditional probability, i.e., the probability of occurring an event y^n given that x^n has already occurred. Using the i.i.d. assumption, we can write this probability as [3]

$$\begin{aligned} p_{Y^n|X^n}(y^n|x^n) &= p_{Y_1|X_1}(y_1|x_1) p_{Y_2|X_2}(y_2|x_2) ... p_{Y_n|X_n}(y_n|x_n) \\ &= p_{Y|X}(y_1|x_1) p_{Y|X}(y_2|x_2) ... p_{Y|X}(y_n|x_n) \\ &= \prod_{i=1}^{n} p_{Y|X}(y_i|x_i). \end{aligned} \tag{2.62}$$

Note that we have two layers of randomness here, one associated with the message chosen by Alice and another one which is associated with the output of the channel.

Now, let $C \equiv \{x^n(m)\}_{m\in[\mathcal{M}]}$ be Alice's coding scheme and $p_e(m, C)$ the total error probability. We can calculate the average error probability of C as

$$\overline{p}_e \equiv \frac{1}{|\mathcal{M}|}\sum_{m=1}^{\mathcal{M}} p_e(m, C). \tag{2.63}$$

In the next section we will see that this probability can be separated into three parts and we will use this fact to demonstrate the channel capacity theorem.

2.4.2 Channel Capacity

Let $I(\mathcal{N})$ be the maximum mutual information associated with the noisy channel $\mathcal{N}$ —the correlation between the input and the output random variables— and $C(\mathcal{N})$ be the channel capacity. The second Shannon's theorem states that [2, 3]

Theorem 2. *The maximum mutual information, $I(\mathcal{N})$, is equal to the capacity $C(\mathcal{N})$ of the channel $\mathcal{N}$.*

If X and Y are random variables associated with the input and output messages, respectively, the mutual information $I(X : Y)$ measures how much information we can get about X once we have received Y. In this way, the maximum information we can get about X is $I(\mathcal{N})$. According to the theorem 2, we have [3]

$$C(\mathcal{N}) = I(\mathcal{N}) \equiv \max I(X : Y). \tag{2.64}$$

The proof of the theorem consists of demonstrating that there is an achievable channel code of type (n, R, ϵ), where n are the number of channel uses, R is the rate of communication and ϵ is some error, which can be made arbitrarily small on the asymptotic limit. Additionally we want to show that the maximum achievable rate $R_{\max}$ is equal to the channel capacity $C(\mathcal{N})$.

To prove the existence of an achievable rate, consider the model code constructed at the beginning of this section. Since there's nothing special about this one we can work with it. Therefore, Alice chooses a message $m \in \mathcal{M}$ which will be codified into a codeword $x^n(m)$ that is selected according to a distribution $p_{X^n}(x^n)$. Then, the noisy channel corrupts the input, so Bob receives the corrupted sequence y^n. This process is shown in Fig. 2.3.

As we did before, we can use the Unit Probability property, i.e., the typical sequences takes all the emission probability as the number of uses of the channel becomes large, and consider that every time Bob receives a sequence y^n that is not on the set of typical sequences, $T_\Delta^{Y^n}$, he declares an error, that we will call $E_0(m)$.

He can also declares other two kinds of errors, $E_1(m)$ and $E_2(m)$. These two are associated with the message sent by Alice, so if m is the message she wants to send, we expect that Bob receives it and nothing different. Thus, he will declares an error, $E_1(m)$, always when he receives a sequence y^n that is in the typical set $T_\Delta^{Y^n}$ but not in the conditionally typical set $T_\Delta^{Y^n|x^n(m)}$ [3]. Moreover, the sequence could be in the typical set $T_\Delta^{Y^n}$, but the received message is some $m' \neq m$, such that the relation [3]

$$\left\{\exists m' \neq m : y^n \in T_\Delta^{Y^n|x^n(m')}\right\} \tag{2.65}$$

is satisfied. In this last case he will declare an error $E_2(m)$.

Therefore, we can write the expectation of the average error probability by

$$\mathbb{E}_C\left\{\overline{p}_e(C)\right\} = \mathbb{E}_C\left\{\frac{1}{|\mathcal{M}|}\sum_m \Pr[E_0(m) \cup E_1(m) \cup E_2(m)]\right\}, \tag{2.66}$$

and, due to the linearity of the expectation value, we can write it as

$$\frac{1}{|\mathcal{M}|}\sum_m \mathbb{E}_C\left\{\Pr[E_0(m) \cup E_1(m) \cup E_2(m)]\right\}. \tag{2.67}$$

As said before, each codeword agrees with the same distribution $p_{X^n}(x^n)$ and is independent of the others. Therefore, the above equation would remain the same for a message m'.

Consequently, we can find the expectation of the error probability for one message instead of finding it for the whole set. So, supposing that $m = 1$, we will have [3]

$$\begin{aligned}\frac{1}{|\mathcal{M}|}\sum_m \mathbb{E}_C\left\{\Pr[E_0(m) \cup E_1(m) \cup E_2(m)]\right\} = \\ \mathbb{E}_C\left\{\Pr[E_0(1) \cup E_1(1) \cup E_2(1)]\right\}\end{aligned} \tag{2.68}$$

and, by using the set properties, we can also write

$$\begin{aligned} &\mathbb{E}_C\left\{\Pr[E_0(1)\cup E_1(1)\cup E_2(1)]\right\} \leq \\ &\mathbb{E}_C\{\Pr[E_0(1)]\}+\mathbb{E}_C\{\Pr[E_1(1)]\}+\mathbb{E}_C\{\Pr[E_2(1)]\}. \end{aligned} \tag{2.69}$$

Let $I_{T_\Delta^{Y^n}}(y^n)$ be the indicator function, such that

$$I_{T_\Delta^{Y^n}}(y^n)=\begin{cases}0, & if \quad y^n \notin T_\Delta^{Y^n}. \\ 1, & if \quad y^n \in T_\Delta^{Y^n}.\end{cases} \tag{2.70}$$

Then, we can associate the terms [3]

$$1-I_{T_\Delta^{Y^n}}(y^n), \tag{2.71}$$

$$I_{T_\Delta^{Y^n}}(y^n)\left(1-I_{T_\Delta^{Y^n|x^n(m)}}(y^n)\right), \tag{2.72}$$

$$\sum_{m'\neq 1} I_{T_\Delta^{Y^n}}(y^n) I_{T_\Delta^{Y^n|x^n(m')}}(y^n), \tag{2.73}$$

with the occurrence of the error events $E_0(1)$, $E_1(1)$ and $E_2(1)$, respectively [3].

Using Eq. (2.71) in the first expected error probability of Eq. (2.69), we have that

$$\mathbb{E}_C\{\Pr[E_0(1)]\}=\mathbb{E}_{X^n(1)}\left\{\mathbb{E}_{Y^n|X^n(1)}\{1-I_{T_\Delta^{Y^n}}(y^n)\}\right\}, \tag{2.74}$$

where the right hand side of this equation follows because the source selects a message $m=1$, so that a random variable $X^n(1)$ is generated, according to some probability $p_X(x^n)$. Then, assuming that this message was sent and n is sufficiently large, the random variable Y^n is highly likely to be a typical random sequence, i.e., for large n we have that $Y^n \in T_\Delta^{Y^n|X^n}$ with respect to X^n, where Y^n is specified by a conditional probability $p_{Y^n|X^n}$. Additionally, considering the linearity of the expectation value, we can write [3]

$$\begin{aligned} \mathbb{E}_{X^n(1)}\left\{\mathbb{E}_{Y^n|X^n(1)}\{1-I_{T_\Delta^{Y^n}}(y^n)\}\right\} &= \mathbb{E}_{X^n(1)}\left\{\mathbb{E}_{Y^n|X^n(1)}\{1\}\right\} - \\ &- \mathbb{E}_{X^n(1)}\left\{\mathbb{E}_{Y^n|X^n(1)}\{I_{T_\Delta^{Y^n}}\}\right\} \\ &= 1-\mathbb{E}_{X^n(1),Y^n}\left\{I_{T_\Delta^{Y^n}}(y^n)\right\} \\ &= 1-\mathbb{E}_{Y^n}\left\{I_{T_\Delta^{Y^n}}(y^n)\right\} \\ &= \Pr\left\{y^n \notin T_\Delta^{Y^n}\right\} \\ &\leq \epsilon, \end{aligned} \tag{2.75}$$

where we used that $\mathbb{E}\{I_{\mathcal{A}}\} = \Pr\{\mathcal{A}\}$ and $\epsilon \in (0,1)$ [3]. Note that, if y^n is a typical sequence, such that $Y^n \in T_{\Delta}^{Y^n}$, then $I_{T_{\Delta}^{Y^n}}(y^n) = 1$ and the expected value in Eq. (2.74) will be zero, on the other hand, if y^n is not a typical sequence, then it will be equal to 1, so we can say that Eq. (2.74) is the probability of y^n not being a typical sequence, which is less or equal than an error ϵ. Following the same procedure, we can write the second probability on the right hand side of Eq. (2.69) as [3]

$$\begin{aligned}
\mathbb{E}_C\{\Pr[E_1(1)]\} &= \mathbb{E}_{X^n(1)}\left\{\mathbb{E}_{Y^n|X^n(1)}\left\{I_{T_{\Delta}^{Y^n}}(y^n)\left(1 - I_{T_{\Delta}^{Y^n|x^n(1)}}(y^n)\right)\right\}\right\} \\
&\leq \mathbb{E}_{X^n(1)}\left\{\mathbb{E}_{Y^n|X^n(1)}\left(1 - I_{T_{\Delta}^{Y^n|x^n(1)}}(y^n)\right)\right\} \\
&= 1 - \mathbb{E}_{X^n(1)}\left\{\mathbb{E}_{Y^n|X^n(1)}\left(I_{T_{\Delta}^{Y^n|x^n(1)}}(y^n)\right)\right\} \\
&= \mathbb{E}_{X^n(1)}\left\{\Pr_{Y^n|X^n(1)}\left\{y^n \notin T_{\Delta}^{Y^n|x^n(1)}\right\}\right\} \leq \epsilon.
\end{aligned} \tag{2.76}$$

And, lastly, using the Boole's inequality

$$\Pr\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} \Pr(A_i), \tag{2.77}$$

where A_i —for $(i = 1, ..., n)$— is a countable set of events, the final probability of Eq. (2.69) will be

$$\mathbb{E}_C\left\{\Pr[E_2(1)]\right\} \leq \mathbb{E}_C\left\{\sum_{m' \neq 1} I_{T_{\Delta}^{Y^n}}(y^n) I_{T_{\Delta}^{Y^n|x^n(m')}}(y^n)\right\} \tag{2.78}$$

$$= \sum_{m' \neq 1} \mathbb{E}_C\left\{I_{T_{\Delta}^{Y^n}}(y^n) I_{T_{\Delta}^{Y^n|x^n(m')}}(y^n)\right\} \tag{2.79}$$

$$= \sum_{m' \neq 1} \mathbb{E}_{X^n(1),X^n(m'),Y^n}\left\{I_{T_{\Delta}^{Y^n}}(y^n) I_{T_{\Delta}^{Y^n|x^n(m')}}(y^n)\right\} \tag{2.80}$$

which, by taking the expectation value, results

$$\begin{aligned}
\sum_{m' \neq 1} \sum_{x^n(1),x^n(m'),y^n} & p_{X^n}(x^n(1)) p_{X^n}(x^n(m')) p_{Y^n|X^n}(y^n|x^n(1)) \times \\
& I_{T_{\Delta}^{Y^n}}(y^n) I_{T_{\Delta}^{Y^n|x^n(m')}}(y^n).
\end{aligned} \tag{2.81}$$

We can assume that $X^n(1)$ always happens, such that the above equation takes the form

$$\sum_{m' \neq 1} \sum_{x^n(m'),y^n} p_{X^n}(x^n(m')) p_{Y^n}(y^n) I_{T_{\Delta}^{Y^n}}(y^n) I_{T_{\Delta}^{Y^n|x^n(m')}}(y^n). \tag{2.82}$$

Note that $p_{Y^n}(y^n) I_{T_{\Delta}^{Y^n}}(y^n)$ is the probability of the typical set $T_{\Delta}^{Y^n}$. Therefore, for a typical set of size $|T_{\Delta}^{Y^n}| \leq 2^{n[H(Y)+\Delta]}$, as defined in Sec. 2.3, we can use Eq. (2.53) to write

$$p_{Y^n}(y^n) I_{T_{\Delta}^{Y^n}}(y^n) \leq 2^{-n[H(Y)-\Delta]}. \tag{2.83}$$

So, we have

$$\sum_{m'\neq 1}\sum_{x^n(m'),y^n} p_{X^n}(x^n(m'))p_{Y^n}(y^n)I_{T_\Delta^{Y^n}}(y^n)I_{T_\Delta^{Y^n|x^n(m')}}(y^n) \leq \\ 2^{-n[H(Y)-\Delta]}\sum_{m'\neq 1}\sum_{x^n(m'),y^n} p_{X^n}(x^n(m'))I_{T_\Delta^{Y^n|x^n(m')}}(y^n). \tag{2.84}$$

Since $p_{X^n}(x^n(m'))$ does not depend on y^n, the above equation can be written as

$$2^{-n[H(Y)-\Delta]}\sum_{m'\neq 1}\sum_{x^n(m'),y^n} p_{X^n}(x^n(m'))I_{T_\Delta^{Y^n|x^n(m')}}(y^n) = \\ 2^{-n[H(Y)-\Delta]}\sum_{m'\neq 1}\sum_{x^n(m')} p_{X^n}(x^n(m'))\sum_{y^n} I_{T_\Delta^{Y^n|x^n(m')}}(y^n). \tag{2.85}$$

Note that the last sum is equal to $|T_\Delta^{Y^n|x^n(m')}|$. Following the same logic as before, we can write $|T_\Delta^{Y^n|x^n(m')}| \leq 2^{n[H(Y|X)+\Delta]}$. We then obtain [3]

$$2^{-n[H(Y)-\Delta]}\sum_{m'\neq 1}\sum_{x^n(m')} p_{X^n}(x^n(m'))\sum_{y^n} I_{T_\Delta^{Y^n|x^n(m')}}(y^n) \leq \\ 2^{-n[H(Y)-\Delta]}2^{n[H(Y|X)+\Delta]}\sum_{m'\neq 1}\sum_{x^n(m')} p_{X^n}(x^n(m')) \tag{2.86}$$

The sum of the probabilities $p_{X^n}(x^n(m'))$ in $x^n(m')$ must be equal to one. Therefore, since the size of the set of all possible messages is $|\mathcal{M}|$, the sum over $m' \neq 1$ will be

$$\sum_{m'\neq 1} 1 = |\mathcal{M}| - 1, \tag{2.87}$$

where 1 is subtracted of $|\mathcal{M}|$ because we are not considering the message m. Then, the above equation takes the form

$$2^{-n[H(Y)-\Delta]}2^{n[H(Y|X)+\Delta]}\sum_{m'\neq 1}\sum_{x^n(m')} p_{X^n}(x^n(m')) = \\ 2^{-n[H(Y)-\Delta]+n[H(Y|X)+\Delta]}\left(|\mathcal{M}|-1\right), \tag{2.88}$$

and, as $|\mathcal{M}| - 1 \leq |\mathcal{M}|$, we get

$$2^{-n[H(Y)-\Delta]+n[H(Y|X)+\Delta]}\left(|\mathcal{M}|-1\right) \leq |\mathcal{M}|\, 2^{-n[H(Y)-H(Y|X)-2\Delta]}. \tag{2.89}$$

By using the mutual information in Eq. (2.25), we can write the final probability as

$$\mathbb{E}_C\left\{\Pr[E_2(1)]\right\} \leq |\mathcal{M}|\, 2^{-n[I(X:Y)-2\Delta]}. \tag{2.90}$$

Therefore, the expectation of the average error probability of a random selected code will be

$$\begin{aligned}\mathbb{E}_C\left\{\Pr[E_0(1)]\right\}+\mathbb{E}_C\left\{\Pr[E_1(1)]\right\}+\mathbb{E}_C\left\{\Pr[E_2(1)]\right\}\leq\\ \epsilon+\epsilon+|\mathcal{M}|\,2^{-n[I(X:Y)-2\Delta]}.\end{aligned}\tag{2.91}$$

Choosing a message set of size $|\mathcal{M}|=2^{n[I(X:Y)-3\Delta]}$, such that we can make the error arbitrarily small [3], we obtain

$$\begin{aligned}2\epsilon+|\mathcal{M}|2^{-n[I(X:Y)-2\Delta]} &= 2\epsilon+2^{n(I(X:Y)-3\Delta)-n[I(X:Y)-2\Delta]}\\ &= 2\epsilon+2^{-n\Delta}.\end{aligned}\tag{2.92}$$

We now define

$$\epsilon'\equiv 2\epsilon+2^{-n\Delta},\tag{2.93}$$

which is an upper bound on the average error probability

$$\mathbb{E}_C\left\{\overline{p}_e(C)\right\}\leq\epsilon'.\tag{2.94}$$

If we throw out the worse half of the messages, i.e., the ones with the biggest probability of error, such that the number of messages will be reduced by half, leaving us with $2^{n[I(X:Y)-3\Delta-1/n]}$ messages. The rate of this coding scheme will be

$$R=\frac{\log(|\mathcal{M}|)}{n}=I(X:Y)-3\Delta-\frac{1}{n},\tag{2.95}$$

so, the mutual information is

$$R+3\Delta+\frac{1}{n}=I(X:Y).\tag{2.96}$$

Thus, the maximum rate is

$$\begin{aligned}R_{\max} &= \max I(X:Y)-3\Delta-\frac{1}{n}\\ &= C(\mathcal{N})-\Delta',\end{aligned}\tag{2.97}$$

where $\Delta'=3\Delta+1/n$ and $C(\mathcal{N})$ is the channel capacity. Therefore, we have shown that there's an achievable channel code $(n,C(\mathcal{N})-\Delta',\epsilon')$ with rate $C(\mathcal{N})-\Delta'$. Using the asymptotic limit, we can make the error ϵ' arbitrarily small, so [3]

$$\lim_{n\to\infty}\epsilon=\lim_{n\to\infty}\Delta=0,$$

such that, the maximum achievable rate will be

$$R_{\max} = C(\mathcal{N}) \tag{2.98}$$

In this way, we finish the proofs of Shannon's theorems. We will use concepts that were discussed here, mainly from Shannon's second theorem, to write the capacity of the communication channel in globally hyperbolic spacetimes. Now we will be focused in understanding what are these spacetimes and why we need them.

Therefore, in the next chapter we will develop some basic ideas of general relativity, which will allow us to construct the spacetime structure, as well as elaborate a quantum field theory on it. All of these ideas will be essential in Chapter 4, when we calculate the channel capacity and evaluate the system's total energy.

Chapter 3

Quantum Field Theory in Curved Spacetimes

In this chapter we want to present a brief introduction to the spacetime structure, as well as some concepts of quantum field theory in curved spacetimes. It will be essential to understanding the calculations of the channel capacity and the energy cost for transmitting information.

There is no way to talk about general relativity without discuss concepts like space and time. These were hotly debated by scientists when Einstein presented his theory of special relativity and, later on, the notion of gravity as a spacetime curvature, giving rise to the general relativity.

Newtonian mechanics is very well accepted by the scientific community and people in general. It is responsible for describing most of the situations we go through in our daily lives. But, with the advance of electromagnetism, the brilliant scientist James C. Maxwell presented the first theory of fields by describing the electromagnetic phenomena in terms of fields in his article "*A dynamical theory of the electromagnetic field*" [21].

It took some time before physicists realized the real significance of Maxwell's field equations —the ones governing the dynamics of the field. Through these equations it was possible to deduce the equation of an electromagnetic wave, from which a constant $c = 1/\sqrt{\mu_0 \epsilon_0}$ appears [22]. By experimental means, the scientists realized that this is precisely the speed of light in vacuum, thus explaining light as travelling electromagnetic waves.

Therefore, according to electromagnetism, speed of light is a constant, so it must be invariant under coordinate transformations. But if we use Galilean transformations —which is fundamental to describing relative motion on Newtonian mechanics— we achieve different results, i.e., the speed of light is observer dependent in Newtonian mechanics. In this way, it was clear that one of the theories needed to be generalized. Therefore, since the experimental results were proving the constancy of speed of light, Einstein proposed a generalization of Newtonian mechanics with the Special Relativity (SR), which is responsible to change completely what we understood about time and space, giving rise to what we call spacetime.

But the generalization were not completed, since SR did not cover gravity. Thus, later on, he presented the General Relativity, from which he elevated the concept of gravity to another level. Before, we thought of gravity as a field intrinsic to each massive body which, in turn, causes them to attract each other. But, according to GR, gravity is the distortion of spacetime in response to the presence of matter, i.e., gravity is the curvature of spacetime.

In this chapter we focus on giving some basics ideas of quantum field theory (QFT) in curved spacetimes. But, before introducing the quantum field itself, we will provide an introduction of GR so we can understand the background structure. Here, we will discuss its postulates and some basics about tensors, covariant differentiation and the Lorentzian metric.

Then we will use this background and formulate a quantum theory for a Klein-Gordon (KG) scalar field ϕ, describing its action and the KG field equation. Also, we will define what is called globally hyperbolic spacetimes and the dynamical evolution of the KG field on it.

We will also see that there are infinitely many choices of Hilbert spaces $\mathcal{H}$ that could be used when formulating the theory, which are, in general, unitarily inequivalent. Thus, to solve this problem, we will introduce the algebraic approach which will allow us to consider all unitarily inequivalent states on an equal footing [12]. Moreover, unless it is explicitly stated, we will consider $c = \hbar = G = 1$, use the signal convention $s(g) = (- \ + \ + +)$ and also write the coordinates (t, x, y, z) as (x^0, x^1, x^2, x^3).

3.1 Notions of General Relativity

Before we go through the postulates of GR, let us first discuss some of the most important concepts of Special Relativity (SR), which are essential to understanding the general theory.

3.1.1 Special Relativity: Motivation and Concepts

In SR the concepts of time and space are quite different from those presented in classical mechanics. The Newtonian universe is described by a four-dimensional affine space $\mathbb{A}^4$ and time is defined by a linear map $t : \mathbb{A}^4 \to \mathbb{A}$ which, as we said before, is used to define simultaneous hypersurfaces, i.e., if a and b are simultaneous we must have $t(a - b) = 0$ and we can define a hypersurface Σ such that $a, b \in \Sigma$.

We call a point in spacetime an *event* and a path of a particle through spacetime is described by a parameterized curve, in the future time direction, that is called *world line*. These definitions are the same for special relativity. But in Newton's conception, space and time are absolute quantities, meaning that all the observers will get the same results when measuring distances in space and time. Specially, if a and b are simultaneous events to one observer, they will also be simultaneous to any other observer and, consequently, the causal structure of spacetime is described according to Fig 3.1.

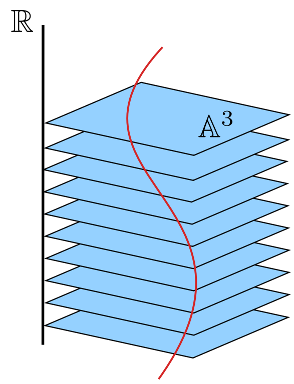

Figure 3.1: Newtonian causal structure of spacetime.

Through the causal structure illustrated by Fig 3.1 we can see that the simultaneous hypersurfaces —the blue surfaces $\mathbb{A}^3$— are crossed by the world-line — the red line— of an observer at only one point. A world-line cannot cross the surface A^3 more than once, in order to do that the observer would need to go back in time which is impossible since we are considering a theory based on deterministic laws, meaning that if the initial conditions of a system are known, then the future state of the system can be predicted with complete accuracy. The hypersurfaces A^3 are called Cauchy surfaces. If a spacetime admits a Cauchy surface it will be known as a globally hyperbolic spacetime.

Although, we have a problem in this spacetime structure, since it admits an absolute conception of space and time, suggesting an instantaneous interaction between bodies [23]. As we saw before, with the advance of electromagnetism it was possible to deduce the equation that governs the dynamics of a electromagnetic field, the wave equation, and was proved experimentally that a electromagnetic wave travel at a constant speed in vacuum, which is the speed of light. According to Galilean transformations [24], it would not be possible, so we have a disagreement between Newtonian mechanics and electromagnetism.

This problem was solved when Einstein presented the Special Relativity (SR) theory [8], which he imposed the following postulates:

1. *Universality of the speed of light*: The speed of light in vacuum is a constant, denoted by c, independent of the state of motion of the source.

2. *Principle of Relativity*: The laws of physics are the same in all inertial frames.

As a consequence of these postulates, Einstein realized that it would be better to replace the absolute concepts of space and time, proposed by Newton, with a four-dimensional structure called spacetime. In this way, coordinate systems must be defined separately to each observer. Additionally, it implies that the idea of absolute simultaneity is extinguished, now we have a simultaneity which depends on the observer.

Due to the constancy of speed of light the causal structure of spacetime must change as well, since light is composed by photons and these are massless particles. Therefore, the speed of light is an upper limit that Einstein needed to take into account. The causal structure which fulfill this requirement is illustrated by the light-cone in Fig. 3.2.

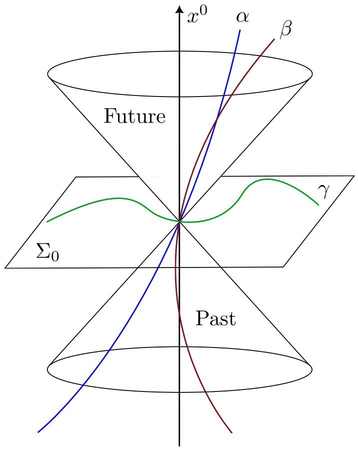

Figure 3.2: A light-cone oriented by the x^0 axis. The red and blue lines are timelike word lines, since they are in the light-cone causal region. But the green one is out of this region, so we call it spacelike world line. Also, note that, the region Σ_0 is a spatial hypersurface where all points on it are simultaneous, relative to x^0, and no timelike world line can pass through it more than one time. Thus, Σ_0 is a Cauchy Surface.

The spacetime is described by a four-dimensional Minkowski space, where space is still Euclidean and time is defined by the real line. This structure is defined by the spacetime metric

$$ds^2 = -(dx^0)^2 + (dx^1)^2 + (dx^2)^2 + (dx^3)^2 = \eta_{\mu\nu} dx^\mu dx^\nu, \tag{3.1}$$

which measures the quadratic distance between two events. Moreover, by investigating Eq. (3.1) as well as Fig. 3.2, note that we have three possibilities for ds^2. When the positive coordinates are greater than dx^0 we have $ds^2 > 0$, it describes the region outside the light-cone, the events within it are called *spacelike* and they are not causally related to the observer in question, i.e., an observer cannot has any kind of influence on spacelike events, the contrary is also true.

A second case can be when dx^0 is greater than the other coordinates, if it happens we have $ds^2 < 0$, which describes the region inside the light-cone. An event in this region is called *timelike*, since the its time coordinate is greater than the spatial ones. These events are causally related to the observer, i.e., an observer can influence and be influenced by any timelike event.

Now, we can also have the situation where the time coordinate and the spatial ones are equal, thus $ds^2 = 0$. For this case, it describes the lines of the light-cone, which are called *null lines*, this region is only allowed to massless particles, it defines the causal relation of spacetime for each observer separately. It means that, a causal relation is observer dependent as well as simultaneity.

But the problem is not completely solved. Special relativity is a theory for the structure of spacetime but is does not include gravity. The general form o the theory was presented by Einstein in 1915 [9]. Here we will present only some basic concepts about this theory, since we are interested in investigating the evolution of a scalar field in this background.

3.1.2 Postulates of General Relativity

We begin our brief discussion by stating three of the four postulates of GR. In order to present the last one, we will first introduce some mathematical approach, since it will be used to describe the postulate. These are:

1. *The Equivalence principle*: Locally, the laws of physics must agree with Special Relativity.

2. *Spacetime is a differential manifold*: As a consequence of the equivalence principle, spacetime is defined as a four-dimensional manifold, $\mathbb{M}$, with a Lorentzian metric $g_{\mu\nu}$, where $\mu, \nu = (0, 1, 2, 3)$, defined on it.

3. *The covariance principle*: The laws of physics must be the same in all reference frames.

The first postulate is important because, since we are generalizing special relativity, we need to get this one from the general theory under the appropriate limit. In GR we see that the presence of gravity implies a curved spacetime. For sufficiently small regions, where gravity can be compared to an uniform acceleration, the structure of spacetime looks like the Minkowski one.

Because of the equivalence principle, Einstein had no other choice than to consider the spacetime structure as a differential manifold. Differential because the laws of physics are written in terms of derivatives. We can define a manifold as a set of smoothly connected open subsets, whose points can be mapped into the real space $\mathbb{R}$. A set is said to be open if, and only if, every point on it has a neighborhood lying in the set. For example, an open set of radius $\mathbf{r}$ and centered at $\mathbf{y}$ is the set of all points $\mathbf{x}$ such that $|\mathbf{x} - \mathbf{y}| < \mathbf{r}$ [23, 25].

Consider an n-dimensional manifold $\mathbb{M}$ where, for each value of n, there is an one-to-one correspondence between an open subset $\mathbb{O}_n \in \mathbb{M}$ and an open subset of $\mathbb{R}^n$, through the mapping [23, 25]

$$\psi_n : \mathbb{O}_n \to \mathbb{E}_n. \tag{3.2}$$

The number of maps n represent the dimension of the manifold and the functions ψ_n along with the open subsets $\mathbb{O}_n$ are known in physics as *coordinate systems*. It can be illustrated by Fig 3.3.

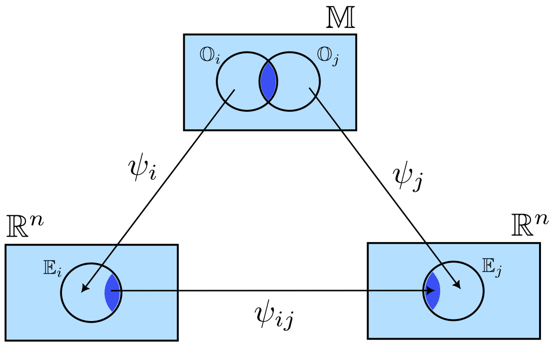

Figure 3.3: Representation of the maps ψ_i, ψ_j and ψ_{ij}.

The entire manifold is covered by the subsets $\mathbb{O}_n$, and these are said to be "smoothly" joined together. In other words, every map ψ_n must be continuous and infinitely differentiable everywhere, since physical equations are written in terms of derivatives — see Appendix A for more details about manifolds.

According to the third postulate of GR, we also need to write the physical equations in such a way that it will be invariant under any coordinate transformation. In the next section we will present an essential tool, which we call tensors, that will allow us to write these equations so that they remain the same in any coordinate system.

3.1.3 Tensors, Metric and Covariant Differentiation

In order to write our equations by using tensors, first we need to understand what are these things and what we can do with them. Here we will make a brief introduction to it and also explain some basics of covariant derivatives, which will be used to write our equations from now on.

Tensors are multi-linear objects that maps a certain number of vectors and dual vectors to a scalar [26]. Suppose that **T** is a tensor of type (*j*, *k*), we say that it maps j dual vectors, V^*, and k vectors, V, to a scalar in $\mathbb{R}$, as [27]

$$T: \bigotimes^{j} V^* \bigotimes^{k} V \to \mathbb{R}. \tag{3.3}$$

For example, a tensor, **p**, of type (2, 1) maps two dual vectors and a vector to a scalar, $p : V^* \times V^* \times V \to \mathbb{R}$. We call (0, 1)-type tensors as dual vectors or, according to some books, one-forms, while a tensor of type (1, 0) is a vector [27]. When specifying the tensor type, as we did in (*j*, *k*), the left number represents how much contravariant slots it has, while the right one constitutes the covariant slots.

For example, as we will see, the solution of the Einstein equations is the metric tensor **g**. It is a (0, 2)-type tensor which is smooth everywhere and also symmetric, i.e., $g_{\mu\nu} = g_{\nu\mu}$. According to Einstein summation, we can use the metric tensor, in some coordinate system, to rise or lower an index of a tensor. For example, we can lower the index of the one-form x^μ by doing

$$x_\nu = g_{\mu\nu} x^\mu, \tag{3.4}$$

where x_ν is the resulting tensor, with $\mu, \nu = (0, 1, 2, 3)$.

It is important to note that, the slots filled with certain indices must be respected throughout all the equation, i.e., all the equation must contain tensors of the same type. Therefore, if we write an equation with a tensor of type (3, 1), which has three contravariant slots filled and one covariant slot filled, as [25]

$$T^{\mu\nu\gamma}_{\alpha} = S^{\mu\nu} \otimes K^{\gamma}_{\alpha}, \tag{3.5}$$

where, again, $\mu, \nu, \gamma, \alpha = (0, 1, 2, 3)$, then the tensors of the right side of the equation must also be of type (3, 1). We can also write

$$T^{\mu\nu\gamma}_{\alpha} = T^{\mu\nu}_{\alpha\beta} g^{\gamma\beta} = T^{\mu\nu\beta\gamma}_{\alpha\beta}. \tag{3.6}$$

But note that, in this case, the index β is being contracted, and $g^{\gamma\beta}$ is the inverse of the metric, which can be calculated because $g_{\mu\nu}$ is not degenerate, i.e., $g = det(g_{\mu\nu}) \neq 0$.

Also, the metric tensor, **g**, determines the smallest distance between two points in spacetime. The quadratic of this distance, and the general form of Eq. (3.1), can be written as

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu, \tag{3.7}$$

where dx^μ and dx^ν are one-forms and the whole right side of the equation, $g_{\mu\nu} dx^\mu dx^\nu$, is the scalar product between the one-forms.

Moreover, the metric tensor has much more important implications than allow us calculate path lengths and scalar products. It also determines the causality, such that no massive particle can reach the speed of light, and, consequently, it gives us a notion of past and future, which we could represent locally throughout light-cones. But in large scale, i.e., considering the endpoints of light world-line, we can imagine the causal structure of GR in a different way.

In order to understand the causal structure of GR, let $S \subset \mathbb{M}$ be an achronal subset, i.e., a set of events or points in spacetime that cannot be causally connected. In other words, it is a set of events that are not related by cause and effect and cannot influence each other, even indirectly. Its domain dependence is $D(S) = D^+ \cup D^-$. Where, D^+ is the set of all points which can be influenced by an event in S and D^- is the set of all points which can influence an event in S. The boundaries of D^+ and D^- are given by the Cauchy horizons h^+ and h^-, respectively, which are the endpoints of light world-lines. The causal structure of a subset $S \subset \mathbb{M}$, which satisfies the above conditions, can be illustrated by Fig. 3.4.

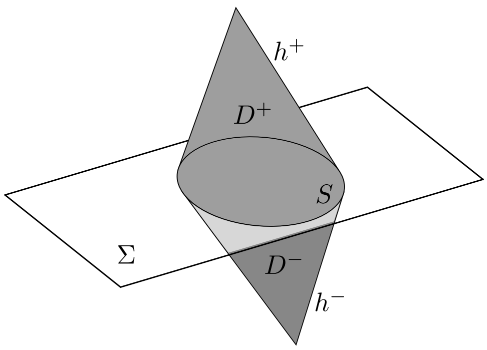

Figure 3.4: Causal relation of a subset $S \subset \mathbb{M}$.

The signature of a metric is the number of negative and positive eigenvalues. The Minkowski metric in Eq. (3.1), for example, has signature $(-,+,+,+)$ or, in some cases, $(+,-,-,-)$, but, as we said before, here we adopt the first one. The signature remains the same for every point in spacetime, since it is continuous and the metric is non-degenerate. Also, because of the minus sign of the signature, we call it a *Lorentzian metric*. In GR we are interested in Lorentzian metrics, since we need it to satisfy the equivalence principle.

In GR framework we cannot make derivatives as we usually do in Newtonian structure. We need to take into account the spacetime curvature, so the derivatives must depends on the metric tensor somehow. We do that by considering the variation of the basis vectors as well, which is expected, since we are moving a vector in a curved spacetime. Therefore, suppose a vector $\vec{V}(x^\alpha)$, and its variation along a curve x^α, in the μ direction, with $\alpha, \mu = (0,1,2,3)$, being defined by the gradient

$$\boldsymbol{\nabla}_\mu \vec{V} = \frac{\partial V^\alpha}{\partial x^\mu} \vec{e_\alpha}. \tag{3.8}$$

$\vec{e}_\alpha$ are the basis vectors of $\vec{V}$. But since we are considering a curved spacetime, the basis vectors do not remain constant when we derive a vector field, and this variation is not being taken into account in the case of the above equation. To do so, we need to add this change to the gradient expression. Thus, we will use an operator that takes into account the neighborhood of an arbitrary point in the manifold and, at the same time, it is reduced to a partial derivative in the case where the manifold is completely flat, such that the equivalence principle is satisfied.

Now, taking into account the variation of the basis vectors, we will operate the gradient $\boldsymbol{\nabla}_\mu$ on the vector $\vec{V} = V^\alpha \vec{e}_\alpha$, such that

$$\boldsymbol{\nabla}_\mu \vec{V} = \frac{\partial \vec{V}}{\partial x^\mu} = \frac{\partial V^\alpha}{\partial x^\mu} \vec{e}_\alpha + V^\alpha \frac{\partial \vec{e}_\alpha}{\partial x^\mu}. \tag{3.9}$$

The variation of basis vectors can be defined as

$$\frac{\partial \vec{e}_\alpha}{\partial x^\mu} \equiv \Gamma^\nu_{\mu\alpha} \vec{e}_\nu. \tag{3.10}$$

The coefficients $\Gamma^\alpha_{\mu\nu}$ is called connection coefficients. It describes how much the basis vectors change throughout a given coordinate system.

Substituting Eq. (3.10) into Eq. (3.9) we obtain

$$\boldsymbol{\nabla}_{\mu}\vec{V} = \frac{\partial V^{\alpha}}{\partial x^{\mu}}\vec{e}_{\alpha} + V^{\alpha}\Gamma^{\nu}_{\mu\alpha}\vec{e}_{\nu}. \tag{3.11}$$

Therefore, the covariant derivative of the components V^{α} of the vector $\vec{V}$ is

$$\boldsymbol{\nabla}_{\mu}V^{\alpha} = \frac{\partial V^{\alpha}}{\partial x^{\mu}} + V^{\nu}\Gamma^{\alpha}_{\mu\nu}, \tag{3.12}$$

where the first part on the right hand side of the equation represents a partial derivative of the field in the direction μ, while the second part is a correction to make the result of the operation covariant, i.e, covariant derivatives are frame independent and hold in any coordinate basis.

In General Relativity, if we consider that the spacetime is curved and torsion free, we work with a unique connection that is said to be metric compatible. A connection is metric compatible if it satisfies the relation $\boldsymbol{\nabla}_{\rho}g_{\mu\nu} = 0$. This implies that the connection can be expressed in terms of the metric as

$$\Gamma^{\sigma}_{\mu\nu} = \frac{1}{2}g^{\sigma\rho}(\partial_{\mu}g_{\nu\rho} + \partial_{\nu}g_{\rho\mu} - \partial_{\rho}g_{\mu\nu}) \quad \rho, \sigma, \mu, \nu = (0, 1, 2, 3). \tag{3.13}$$

This unique connection is usually called *Christoffel symbols* and it is very important for calculations in GR framework, since it appears in many equations in GR.

Now that we saw some basics of tensors and covariant derivatives, and also the metric tensor, we are ready to present the last postulate of GR. Since we will not work with it in this article, the last postulate will be introduced in a matter of completeness. It is a statement about the dynamics of the field, the Einstein field equations [23, 25, 28]

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = 8\pi G T_{\mu\nu}. \tag{3.14}$$

$R_{\mu\nu}$ is known as the Ricci tensor, which can be calculated by contracting the Riemann curvature tensor

$$R^{\alpha}_{\mu\nu\sigma} = \left(\frac{\partial\Gamma^{\alpha}_{\sigma\mu}}{\partial x^{\nu}} + \Gamma^{\alpha}_{\nu\lambda}\Gamma^{\lambda}_{\sigma\mu} - \frac{\partial\Gamma^{\alpha}_{\nu\mu}}{\partial x^{\sigma}} - \Gamma^{\alpha}_{\sigma\lambda}\Gamma^{\lambda}_{\nu\mu}\right). \tag{3.15}$$

as

$$R_{\mu\nu} \equiv \sum_{\alpha} R^{\alpha}_{\mu\nu\alpha} = R^{0}_{\mu\nu0} + R^{1}_{\mu\nu1} + R^{2}_{\mu\nu2} + R^{3}_{\mu\nu3}. \tag{3.16}$$

The Riemann tensor describes how much two paths, which has the same initial and final points, changes with the curvature of the spacetime. If there is no curvature, the Riemann tensor will vanish, which means that the metric is constant and all the laws of physics will be reduced to those of SR, as it must be, according to the equivalence principle. The tensor $R^{\alpha}_{\mu\nu\sigma}$, which depends on the path, is anti-symmetric in the last two indices, i.e, $R^{\alpha}_{\mu\nu\sigma} = -R^{\alpha}_{\mu\sigma\nu}$. Moreover, since the Riemann tensor is symmetric in μ and ν the Ricci tensor must also be symmetric.

We can do the same to the Ricci tensor in order to obtain the curvature scalar, R, as

$$R \equiv g^{\mu\nu} R_{\mu\nu}. \tag{3.17}$$

Moreover, we can also write the left hand side of Eq. (3.14) by defining the Einstein tensor

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R, \tag{3.18}$$

which is symmetric and conserved, i.e., $\boldsymbol{\nabla}^{\mu} G_{\mu\nu} = 0$.

In the right hand side of Eq. (3.14) we have G, which is Newton's constant, and the stress-energy tensor, $T_{\mu\nu}$, that describes the energy-matter content of spacetime and must also be symmetric and conserved, since the left side of the equation is. Therefore, the left had side of Eq. (3.14) determines how the spacetime responds to the presence of matter, which is distributed according to the stress-energy tensor that is in the right hand side of the equation.

Note that, the Einstein field equation will be solved if we find the components of the metric tensor which satisfies it, since each tensor in the equation depends on them.

With the last postulate of GR being introduced, we finish our discussion about this topic. Now, our task is to introduce a scalar field in this background, such that we can get an wave equation that describes the dynamical evolution of the quantum field on a curved spacetime.

3.2 Wave Equation and Globally Hyperbolic Manifolds

Considering the spacetime structure ($\mathbb{M}$,$g_{\mu\nu}$) constructed above, where $\mathbb{M}$ is a four-dimensional manifold and $g_{\mu\nu}$ is a Lorentzian metric and it is related to the energy-matter content of spacetime by the Einstein field equations (3.14). We can use this background to formulate a quantum theory of a Klein-Gordon scalar field ϕ [12].

Let us consider a scalar real field that is described by the function $\phi(x^\alpha)$, where $\alpha = (0, 1, 2, 3)$. This is similar to the set of generalized coordinates, q_i, used in classical mechanics. But now, we are describing the field on an arbitrary region, $\mathcal{U}$, of the spacetime manifold, bounded by a closed hypersurface $\partial\mathcal{U}$, i.e., there are no points on the hypersurface that are not part of the hypersurface itself. The Lagrangian density, $\mathcal{L}$, will depend on the scalar field ϕ and its first derivatives, $\nabla_\alpha\phi$. Therefore, we can describe the action as [29, 30]

$$S(\phi) = \int_{\mathcal{U}} \mathcal{L}(\phi, \phi_{;\alpha}, \alpha)\sqrt{-g}\, d^4x, \tag{3.19}$$

where $g \equiv \det(g_{\mu\nu})$ and we are using that $\phi_{;\alpha} \equiv \nabla_\alpha\phi$. Introducing the variation $\delta\phi(x^\alpha)$, which satisfies

$$\delta\phi(x^\alpha)\bigg|_{\partial\mathcal{U}} = 0, \tag{3.20}$$

and the variational principle,

$$\delta S = \delta \int_{\mathcal{U}} \mathcal{L}(\phi, \phi_{;\alpha}, \alpha)\sqrt{-g}\, d^4x = 0, \tag{3.21}$$

we get

$$\delta S = \int_{\mathcal{U}} \left(\frac{\partial\mathcal{L}}{\partial\phi}\delta\phi + \frac{\partial\mathcal{L}}{\partial\phi_{;\alpha}}\delta\phi_{;\alpha} \right) \sqrt{-g}\, d^4x \tag{3.22}$$

Using the notation $\mathcal{L}' \equiv \frac{\partial\mathcal{L}}{\partial\phi}$ and $\mathcal{L}^\alpha \equiv \frac{\partial\mathcal{L}}{\partial\phi_{;\alpha}}$, the Eq. (3.22) becomes

$$\delta S = \int_{\mathcal{U}} \left(\mathcal{L}'\delta\phi + \mathcal{L}^\alpha\delta\phi_{;\alpha}\right)\sqrt{-g}\, d^4x. \tag{3.23}$$

The second term on the parentheses can be calculated using integration by parts. Thus,

$$\begin{aligned}
\int_{\mathcal{U}} \mathcal{L}^\alpha\delta\phi_{;\alpha}\sqrt{-g}\, d^4x &= \mathcal{L}^\alpha\delta\phi\bigg|_{\mathcal{U}} - \int_{\mathcal{U}} \nabla_\alpha(\mathcal{L}^\alpha)\delta\phi\sqrt{-g}\, d^4x &(3.24)\\
&= \int_{\mathcal{U}} \left[\nabla_\alpha(\mathcal{L}^\alpha\delta\phi) - \nabla_\alpha(\mathcal{L}^\alpha)\delta\phi\right]\sqrt{-g}\, d^4x. &(3.25)
\end{aligned}$$

In this way, Eq. (3.23) will be

$$\begin{aligned}
\delta S &= \int_{\mathcal{U}} \left[\mathcal{L}'\delta\phi + \nabla_\alpha\mathcal{L}^\alpha\delta\phi - \nabla_\alpha(\mathcal{L}^\alpha)\right]\sqrt{-g}\, d^4x &(3.26)\\
&= \int_{\mathcal{U}} \left(\mathcal{L}' - \nabla_\alpha\mathcal{L}^\alpha\right)\delta\phi\sqrt{-g}\, d^4x + \oint_{\partial U} \mathcal{L}^\alpha\delta\phi d\Sigma_\alpha, &(3.27)
\end{aligned}$$

where, on the last step, we used the Gauss-Stokes theorem, which says that for any vector field $\mathbf{A}^\alpha$ defined within $\mathcal{U}$,

$$\int_{\mathcal{U}} \nabla_\alpha(\mathbf{A}^\alpha)\sqrt{-g}\, d^4x = \oint_{\partial\mathcal{U}} \mathbf{A}^\alpha\, d\Sigma_\alpha, \tag{3.28}$$

with $d\Sigma_\alpha$ being the surface element [30]. Now, due to Eq. (3.20), we can write δS as

$$\delta S = \int_{\mathcal{U}} (\mathcal{L}' - \nabla_\alpha \mathcal{L}^\alpha)\, \delta\phi \sqrt{-g}\, d^4x = 0, \tag{3.29}$$

thus,

$$\nabla_\alpha \mathcal{L}^\alpha - \mathcal{L}' = 0, \tag{3.30}$$

or

$$\nabla_\alpha \left(\frac{\partial \mathcal{L}}{\partial \phi_{;\alpha}} \right) - \frac{\partial \mathcal{L}}{\partial \phi} = 0, \tag{3.31}$$

which is known as the Euler-Lagrange equation for a scalar field ϕ.

As suggested by Eq. (3.31) the Lagrangian density, $\mathcal{L}$, must depends only on ϕ and its first derivatives. Also, since the wave equation is written in terms of second derivatives, the density $\mathcal{L}$ should be a quadratic function in ϕ and its derivatives [29]. Considering that ϕ is the Klein-Gordon field, with Lagrangian density given by

$$\mathcal{L} = -\frac{1}{2}\sqrt{-g}\left(g^{\mu\nu}\nabla_\mu\phi\nabla_\nu\phi + m^2\phi^2\right), \tag{3.32}$$

where m is the field mass and $g^{\mu\nu}$ is the inverse of $g_{\mu\nu}$. Consequently, the action of a Klein-Gordon field in a curved spacetime is

$$S = -\frac{1}{2}\int_{\mathbb{M}} d^4\mathbf{x}\sqrt{-g}\,(\nabla_\mu\nabla^\mu\phi + m^2\phi^2). \tag{3.33}$$

Now we can use the Eq. (3.31) along with Eq. (3.32) to derive the equation of motion, as

$$\frac{\partial \mathcal{L}}{\partial(\nabla_\mu\phi)} = -\sqrt{-g}\, g^{\mu\nu}\nabla_\mu\phi, \tag{3.34}$$

and

$$\frac{\partial \mathcal{L}}{\partial \phi} = -\sqrt{-g}\, m^2\phi. \tag{3.35}$$

Therefore, we have

$$\nabla^{\mu}\nabla_{\mu}\phi - m^2\phi = 0, \tag{3.36}$$

which is known as the *Klein-Gordon equation.*

In order to actually describe the dynamics of the system, Eq. (3.36) needs to satisfy a well posed initial value problem so that we can achieve unique solutions. According to the postulates of General Relativity, for each point in the four-dimensional manifold $\mathbb{M}$ we can construct light-cones which are time oriented, i.e., each one of it has a well defined future and past, as illustrated in Fig. 3.2. Consequently, there will be a three-dimensional hypersurface Σ_0, such that, the points on it are simultaneous to each other. This will allow us to formulate a well posed initial value theory in the sense that, we can consider points in a hypersurface as the one described above, such that we will have a well defined time orientation.

Therefore, consider the set of hypersurfaces Σ_t for which every pair of points $p, q \in \Sigma_t$ is simultaneous. If $\Sigma_t \in \mathbb{M}$, we say that this hypersurface is a submanifold of $\mathbb{M}$, i.e., a three-dimensional region in $\mathbb{M}$ where every timelike curve on the manifold may be extended to a timelike curve that intersects Σ_t in one point.

Hypersurfaces with this restriction are called *Cauchy surfaces*, and a spacetime which admits a Cauchy surface is called *Globally Hyperbolic Spacetime.* Now, we can expect a deterministic dynamical evolution with well defined initial values in Σ. We can state the following theorem [12].

Theorem 3. *If ($\mathbb{M}$, $g_{\mu\nu}$) is a globally hyperbolic spacetime with smooth Cauchy surface Σ, then the dynamical evolution of the Klein-Gordon field can be specified by a pair of smooth functions (ϕ_0, $\dot{\phi}_0$) on Σ, where $\dot{\phi}_0 = n^{\mu}\nabla_{\mu}\phi$ and n^{μ} is a unitary vector normal to Σ.*

Therefore, if we are going to work with a globally hyperbolic spacetime we can parameterize Eqs. (3.33) and (3.32) in terms of a time flow vector $t^{\mu} = Nn^{\mu}$, such that $t^{\mu}\boldsymbol{\nabla}_{\mu}t = 1$, where n^{μ} is an unitary vector normal to a hypersurface Σ_t of constant t and N is a lapse function. Thus, we can rewrite the Lagrangian density (3.32) as

$$\mathcal{L} = -\frac{1}{2}\sqrt{-g}\left[(n^{\mu}\boldsymbol{\nabla}_{\mu}\phi)^2 + h^{\mu\nu}\nabla_{\mu}\phi\nabla_{\nu}\phi + m^2\phi^2\right]N, \tag{3.37}$$

where $h^{\mu\nu}$ is the metric induced by g on Σ_t. Also, the action becomes

$$S = -\frac{1}{2}\int_{\mathbb{M}} \sqrt{-g}\left[(n^{\mu}\boldsymbol{\nabla}_{\mu}\phi)^2 + h^{\mu\nu}\nabla_{\mu}\phi\nabla_{\nu}\phi + m^2\phi^2\right] N d^4\mathbf{x}. \tag{3.38}$$

We can use the relation

$$n^{\mu}\boldsymbol{\nabla}_{\mu}\phi = \frac{1}{N}t^{\mu}\boldsymbol{\nabla}_{\mu}\phi = \frac{1}{N}\dot{\phi}, \tag{3.39}$$

to calculate the momentum density as

$$\begin{aligned}\pi &= \frac{\partial\mathcal{L}}{\partial\dot{\phi}} = \frac{1}{2}N\sqrt{-g}\frac{\partial}{\partial\dot{\phi}}(n^{\mu}\boldsymbol{\nabla}_{\mu}\phi)^2 = \frac{1}{2}N\sqrt{-g}\frac{\partial}{\partial\dot{\phi}}\left(\frac{\dot{\phi}}{N}\right)^2 \\ &= \sqrt{-g}\left(\frac{\dot{\phi}}{N}\right) = (n^{\mu}\boldsymbol{\nabla}_{\mu}\phi)\sqrt{-g}, \end{aligned} \tag{3.40}$$

which is canonically conjugate to ϕ, such that, according to the Hamiltonian formulation studied in classical mechanics, we can write the canonical Hamiltonian of the field as [12]

$$H_{\phi}(t) \equiv \int_{\Sigma_t}\left[\pi(t,\mathbf{x})\dot{\phi}(t,\mathbf{x}) - \mathcal{L}(\phi,\boldsymbol{\nabla}_{\mu}\phi)\right] d^3\mathbf{x}. \tag{3.41}$$

Following the canonical quantization procedure, since the operators ϕ and π are canonical conjugate quantities, they must satisfy the equal time commutation relations [11, 12]

$$[\phi(t,\mathbf{x}),\phi(t,\mathbf{x'})]_{\Sigma_t} = [\pi(t,\mathbf{x}),\pi(t,\mathbf{x'})]_{\Sigma_t} = 0, \tag{3.42}$$

$$[\phi(t,\mathbf{x}),\pi(t,\mathbf{x'})]_{\Sigma_t} = i\delta^3(\mathbf{x},\mathbf{x'}), \tag{3.43}$$

where $\mathbf{x} \equiv (x^1, x^2, x^3)$ are spatial coordinates defined on Σ_t. The state of a system at any instant of time is described by a point in phase space, $\mathcal{M}$, which can be determined by the identification of smooth functions $\phi(t,\mathbf{x})$ and $\pi(t,\mathbf{x})$ on a Cauchy surface Σ_t. Then, the phase space $\mathcal{M}$ shall be defined as [12]

$$\mathcal{M} \equiv \left\{\phi : \Sigma_t \to \mathbb{C}, \pi : \Sigma_t \to \mathbb{C} \mid \phi, \pi \in C_0^{\infty}(\Sigma_t)\right\}, \tag{3.44}$$

where $\mathbb{C}$ is the set of complex numbers and $C^{\infty}(\Sigma_t)$ is the set of infinitely differentiable compact support functions in the Cauchy surface Σ_t.

Hence, with these restrictions about the spacetime structure, we formulate a field theory in which the Klein-Gordon equation (3.36) satisfies well posed initial values. Additionally, we must have a solution space to Eq. (3.36), since it is possible to solve it now. Thus, in a globally hyperbolic spacetime, we can define Cauchy surfaces which, for any of them, we specify smooth functions ϕ and π, such that the pair $[\phi, \pi]$ represents a point in phase space, $\mathcal{M}$, which in turn, leads to a unique element of $\mathcal{S}^{\mathbb{C}}$, where $\mathcal{S}^{\mathbb{C}}$ is the space of complex solutions of Eq. (3.36).

Now, if we have two points in $\mathcal{M}$, say $[\phi_1, \pi_1]$ and $[\phi_2, \pi_2]$, which gives rise to two elements of $\mathcal{S}^{\mathbb{C}}$, we can define a bilinear map, Ω, that acts on the space $\mathcal{S}^{\mathbb{C}}$ as

$$\Omega : \mathcal{S}^{\mathbb{C}} \times \mathcal{S}^{\mathbb{C}} \to \mathbb{C}. \tag{3.45}$$

Ω is a symplectic structure that defines a fundamental observable in the phase space $\mathcal{M}$. Independently of the choice of coordinates, we can write it as

$$\Omega\left([\phi_1, \pi_1], [\phi_2, \pi_2]\right) \equiv \int_{\Sigma_t} \left(\pi_1 \phi_2 - \pi_2 \phi_1\right) d^3\mathbf{x}, \tag{3.46}$$

with the volume element in Σ_t given by $d^3\mathbf{x} \equiv dx^1 \wedge dx^2 \wedge dx^3$. Using Eq. (3.40) we get

$$\Omega\left(\phi_1, \phi_2\right) \equiv \int_{\Sigma_t} \sqrt{-g}\, n^\mu \left[\phi_2 \boldsymbol{\nabla}_\mu \phi_1 - \phi_1 \boldsymbol{\nabla}_\mu \phi_2\right] d^3\mathbf{x}. \tag{3.47}$$

So that we can define the Klein-Gordon inner product of the functions ϕ_1 and ϕ_2, as [11–13]

$$\langle \phi_1 | \phi_2 \rangle_{\text{KG}} \equiv -i\, \Omega(\overline{\phi}_1, \phi_2). \tag{3.48}$$

Thus

$$\langle \phi_1 | \phi_2 \rangle_{\text{KG}} = -i \int_{\Sigma_t} \sqrt{-g}\, n^\mu \left[\phi_2 \boldsymbol{\nabla}_\mu \overline{\phi}_1 - \overline{\phi}_1 \boldsymbol{\nabla}_\mu \phi_2\right] d^3\mathbf{x}, \tag{3.49}$$

where the bar stands for complex conjugation. Also, it must satisfies all the properties of an inner product

i. if V is a vector space, then the inner product on V is a map $\langle | \rangle : V \times V \to \mathbb{C}$, with $\mathbb{C}$ being the set of complex numbers;

ii. if $\psi_1, \psi_2 \in V$ the inner product must satisfy $\overline{\langle \psi_1 | \psi_2 \rangle} = \langle \psi_2 | \psi_1 \rangle$;

iii. if $\psi \in V$ then $\langle \psi | \psi \rangle > 0$ for all $\psi \neq 0$;

iv. if $\langle|\rangle$ is an inner product on the vector space V and $\psi \in V$, then we can define the norm on it by $||\psi|| \equiv (\langle\psi|\psi\rangle)^{1/2}$.

With the inner product defined by Eq. (3.48), we can choose a subspace $\mathcal{H} \subset \mathcal{S}^{\mathbb{C}}$, such that $\langle\,|\,\rangle_{\mathrm{KG}}$ is positive on it. Also, let $u \in \mathcal{H}$ and $v \in \overline{\mathcal{H}}$ being basis vectors, where $\overline{\mathcal{H}}$ is the dual subspace, such that $\mathcal{S}^{\mathbb{C}} = \mathcal{H} \oplus \overline{\mathcal{H}}$. If the inner product satisfies the relation $\langle u|v\rangle_{\mathrm{KG}} = 0$ and it is positive and we can define $(\mathcal{H}, \langle\,|\,\rangle_{\mathrm{KG}})$ as a Hilbert space with the orthonormal basis $\{u_j\}$.

In quantum mechanics, we can write the observable Ω in terms of the creation and annihilation operators. For that, we construct the symmetric Fock space $\mathcal{F}_S(\mathcal{H})$ based upon $\mathcal{H}$ [11, 12, 27]

$$\mathcal{F}_S(\mathcal{H}) \equiv \bigoplus_{n=0}^{\infty} S_\nu H^{\oplus n}, \tag{3.50}$$

where S_ν is the operator that symmetrizes or antisymmetrizes a tensor. If the Hilbert space describes particles obeying bosonic $(\nu = +)$ statistics then, the Fock space will be the sum of symmetrized tensor products and if it describes particles obeying fermionic $(\nu = -)$ statistics the sum will be over antisymmetrized tensor products. We can use the orthonormal basis $\{u_j\}$ to make an expansion of the quantum field operator as [11]

$$\hat{\phi}(t, \mathbf{x}) \equiv \sum_j \left[u_j(t, \mathbf{x}) a(\overline{u}_j) + \bar{u}_j(t, \mathbf{x}) a^\dagger(u_j)\right], \tag{3.51}$$

where, $a(\bar{u}_j)$ and $a^\dagger(u_j)$ are the creation and annihilation operators, respectively. They satisfy the relations

$$[a(\bar{u}_j), a^\dagger(u_j)] = \langle u|v\rangle\, \mathbb{I}, \tag{3.52}$$

where $\mathbb{I}$ is the identity operator in $\mathcal{F}_S(\mathcal{H})$. Finally, the fundamental quantum observables $\hat{\Omega}(\psi, \cdot)$ on $\mathcal{F}_S(\mathcal{H})$ can be defined in terms of the annihilation and creation operators, for each $\psi \in \mathcal{S}^{\mathbb{C}}$, as [11–13]

$$\hat{\Omega}(\psi, \cdot) \equiv ia(\overline{K\psi}) - ia^\dagger(K\psi), \tag{3.53}$$

where K is a linear one-to-one map $K : \mathcal{S}^{\mathbb{C}} \to \mathcal{H}$, which takes an element of the space of complex solutions into an element of the Hilbert space.

The symplectic vector space of complex solutions to Eq. (3.36) is then defined as $(\mathcal{S}^{\mathbb{C}}, \hat{\Omega})$ with the initial data $[\phi, \pi]$. The observable $\hat{\Omega}$ can also be rewritten as an operator-valued distribution, i.e., an operator representing the value of the field. In order to do that, consider the vector space, $\mathcal{T} = C_0^{\infty}(\mathbb{M})$, of smooth, compact-support real functions on the manifold $\mathbb{M}$. If we add some source term $f \in \mathcal{T}$ in the right hand side of Eq. (3.36) we will, then, have unique advanced and retarded solutions to it [12]. Thus, if Af and Rf are the advanced and retarded solutions, respectively, we have [12]

$$(\nabla^{\mu}\nabla_{\mu} - m^2)(Af) \quad = \quad f, \tag{3.54}$$

$$(\nabla^{\mu}\nabla_{\mu} - m^2)(Rf) \quad = \quad f, \tag{3.55}$$

then, the solution to the homogeneous equation (3.36) will be

$$Ef = Af - Rf, \tag{3.56}$$

where $E : \mathcal{T} \to \mathcal{S}$, with $\mathcal{S} \subset \mathcal{S}^{\mathbb{C}}$ being the set of real solutions to Eq. (3.36).

The function E also satisfies the properties [12]

i. For every element $\psi \in \mathcal{S}$ there is an Ef, such that $Ef = \psi$;

ii. $Ef = 0$ in, and only if, $f = (\nabla^{\mu}\nabla_{\mu} - m^2)g$, for $g \in \mathcal{T}$;

iii. For all $\psi \in \mathcal{S}$ and $f \in \mathcal{T}$, we can write

$$\int_{\mathbb{M}} \sqrt{-g}\psi f d^4\mathbf{x} = \Omega(Ef, \psi), \tag{3.57}$$

where $\Omega(Ef, \psi)$ is the function of fundamental observables in $\mathcal{M}$.

In order to prove the first property, consider a real solution of the Klein-Gordon field equation (3.36), $\psi \in \mathcal{S}$, and a smooth function α, such that $\alpha = 0$ to $x^0 \leq 0$ and $\alpha = 1$ to $x^0 \geq 1$. Therefore, if

$$\nabla^{\mu}\nabla_{\mu}\psi - m^2\psi = 0 \tag{3.58}$$

then, we can define a source term $f \in \mathcal{T}$ as

$$f \equiv \left(\nabla^{\mu}\nabla_{\mu} - m^2\right)(-\alpha\psi), \tag{3.59}$$

where $(-\alpha\psi) = Rf$ is a retarded solution, as we can see from Eq. (3.55). Another solution to the in-homogeneous equation is $(1-\alpha)\psi$, thus [12]

$$
\begin{aligned}
(\nabla^\mu\nabla_\mu - m^2)(1-\alpha)\psi &= \nabla^\mu\nabla_\mu\psi - m^2\psi - \nabla^\mu\nabla_\mu(\alpha\psi) + m^2(\alpha\psi) && (3.60)\\
(\nabla^\mu\nabla_\mu - m^2)(-\alpha\psi) &= f && (3.61)
\end{aligned}
$$

so that, we can define $(1-\alpha)\psi = Af$, which is the advanced solution. Now, if we use Eq. (3.56), we get

$$
Ef = (1-\alpha)\psi - (-\alpha\psi) = \psi. \tag{3.62}
$$

Which proves the first property.

The proof of the second property is straightforward. If $f = (\nabla^\mu\nabla_\mu - m^2)g$, with $g \in \mathcal{T}$, then, by using Eqs. (3.54) and (3.55), we can write

$$
\begin{aligned}
(\nabla^\mu\nabla_\mu - m^2)Af &= f = (\nabla^\mu\nabla_\mu - m^2)g && (3.63)\\
(\nabla^\mu\nabla_\mu - m^2)Rf &= f = (\nabla^\mu\nabla_\mu - m^2)g, && (3.64)
\end{aligned}
$$

therefore, $Af = Rf = g$ [12] and, consequently, $Ef = 0$ [12]. Now, if $Ef = 0$, then $Af = Rf = g \in \mathcal{T}$.

To prove the third property, consider the solution to the homogeneous equation $\psi \in \mathcal{S}$, the source function $f \in \mathcal{T}$ and the times $t_1, t_2 \in \mathbb{R}$, with $f = 0 \, \forall t \notin [t_1, t_2]$. Thus, [12]

$$
\int_{\mathbb{M}} \sqrt{-g}\,\psi f d^4\mathbf{x} = \int_{t\in[t_1,t_2]} d^4\mathbf{x}\sqrt{-g}\,\psi(\nabla^\mu\nabla_\mu - m^2)Af. \tag{3.65}
$$

We can use Eq. (3.58) to write

$$
\begin{aligned}
&\int_{t\in[t_1,t_2]} d^4\mathbf{x}\sqrt{-g}\,\psi(\nabla^\mu\nabla_\mu - m^2)Af = \\
&= \int_{t\in[t_1,t_2]} d^4\mathbf{x}\sqrt{-g}\,[\psi(\nabla^\mu\nabla_\mu - m^2)Af - Af(\nabla^\mu\nabla_\mu - m^2)\psi] \\
&= \int_{t\in[t_1,t_2]} d^4\mathbf{x}\sqrt{-g}\,\left[\psi\nabla^\mu\nabla_\mu(Af) - Af\nabla^\mu\nabla_\mu(\psi)\right]. && (3.66)
\end{aligned}
$$

If we use Green's identity,

$$
\int_{\mathbf{V}} (\psi\boldsymbol{\nabla}^2\phi - \phi\boldsymbol{\nabla}^2\psi)d\mathbf{V} = \oint_{\mathbf{S}} (\psi\boldsymbol{\nabla}\phi - \phi\boldsymbol{\nabla}\psi)\cdot d\mathbf{S}, \tag{3.67}
$$

where ψ and ϕ are twice differentiable functions, we get [12]

$$\int_{t=t_1} d^3\mathbf{x}\sqrt{-g}\,[\psi\boldsymbol{\nabla}_t(Ef) - Ef\boldsymbol{\nabla}_t(\psi)] = \Omega(Ef,\psi). \tag{3.68}$$

Since we considered the hypersurface $t = t_1$ then, outside the causal future of f, we have $Rf = 0$, such that $Af = Ef$. Also, the spatial derivatives cancel each other, so we have only time ones. Therefore, we proved all properties of the map E.

Moreover, we can define the field operator associated to a real solution $Ef \in \mathcal{S}$ such as [11, 12]

$$\hat{\phi}(f) \equiv \hat{\Omega}(Ef,\cdot) \equiv ia(\overline{KEf}) - ia^\dagger(KEf), \tag{3.69}$$

with $K : \mathcal{S} \to \mathcal{H}$ being the operator which takes the positive-norm part of any real solution $Ef \in \mathcal{S}$ [11]. Equation (3.69) represents a map from test functions into operators in $\mathcal{F}_S(\mathcal{H})$.

The properties of the function E, which were demonstrated above, has some effects on the field operator $\hat{\phi}(f)$. The first one we already used on the definition of Eq. (3.69), i.e., describing the field operator $\hat{\phi}(f)$ for any $f \in \mathcal{T}$ is much the same as specifying the observable $\Omega(\hat{\psi},\cdot)$ for $\psi \in \mathcal{S}$ [12]. The second property implies that, if $f = (\nabla^\mu\nabla_\mu - m^2)\,g$ for some $g \in \mathcal{T}$, the field operator $\hat{\phi}(f)$, as well as the observable $\hat{\Omega}(Ef,\cdot)$, will be equal to zero. Lastly, we can use the third property to demonstrate that the field operators satisfies the fundamental commutation relations [11]

$$[\hat{\phi}(f),\hat{\phi}(g)] = -i\Delta(f,g)\mathbb{I}, \tag{3.70}$$

for all $f, g \in \mathcal{T}$ and

$$\Delta(f,g) \equiv \int_{\mathbb{M}} d^4\mathbf{x}\sqrt{-g}\,Egf. \tag{3.71}$$

In order to demonstrate it, we can use the definition of Eq. (3.69) to some $\psi \in \mathcal{S}$, and write

$$\left[\hat{\phi}(f),\hat{\phi}(g)\right] = \left[\hat{\Omega}(Ef,\psi),\hat{\Omega}(Eg,\psi)\right]. \tag{3.72}$$

Where, the right hand side of the above equation can be related to the function Ω on the classical phase space $\mathcal{M}$ if we use

$$\left[\hat{\Omega}([\phi_1,\pi_1],\cdot),\hat{\Omega}([\phi_2,\pi_2],\cdot)\right] = -i\Omega([\phi_1,\pi_1],[\phi_2,\pi_2])\mathbb{I}, \tag{3.73}$$

which is another representation of the commutators in Eqs. (3.42) and (3.43).

In the case of equation (3.72) we have [12]

$$\left[\hat{\Omega}(Ef,\psi),\hat{\Omega}(Eg,\psi)\right] = -i\Omega(Ef,Eg), \tag{3.74}$$

such that, we can use the third property of E, defined above, to write

$$\Omega(Ef,Eg) = \int_{\mathbb{M}} d^4\mathbf{x}\sqrt{-g}\,Egf \equiv \Delta(f,g), \tag{3.75}$$

then, the demonstration is finished by

$$\left[\hat{\phi}(f),\hat{\phi}(g)\right] = -i\Omega(Ef,Eg)\mathbb{I} = -i\Delta(f,g)\mathbb{I}. \tag{3.76}$$

The disadvantage of this method is that it involves arbitrary choices of Hilbert spaces, which can take us to unitarily inequivalent representations of the quantum field theory in curved spacetime. Physically, it implies that we could make infinitely choices of particles and vacuum representations, since each Fock space $\mathcal{F}(\mathcal{H})$ represents the particle's states. Therefore, if we have different, i.e., not unitarily equivalent, Hilbert spaces we can define "particles" in distinct ways. We can solve this problem by formulating the theory through the algebraic approach, which will allow us to construct a quantum field theory without making any preferred choice of states.

3.3 The Algebraic Approach

The algebraic formulation of the quantum field theory will grant us that even states derived from unitarily inequivalent representations can be treated equally, since they satisfy the same algebraic relations. To understand it, consider the Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$, which comprises all field states, such that, they can be defined as the symmetric Fock spaces $\mathcal{F}_S(\mathcal{H}_1)$ and $\mathcal{F}_S(\mathcal{H}_2)$, respectively.

As we did before, we can define fundamental quantum observables on each Fock space, $\mathcal{F}_S(\mathcal{H}_1)$ and $\mathcal{F}_S(\mathcal{H}_2)$, as $\hat{\Omega}_1(\psi,\cdot)$ and $\hat{\Omega}_2(\psi,\cdot)$, respectively. Thus, we have two different representations of a field theory. These two theories may be unitarily inequivalent, which would force us to choose one of them as preferred. In order to avoid that choice, we can formulate a quantum field theory such that the relations satisfied by the operators $\{\hat{\Omega}_1(\psi, \cdot)\} \in \mathcal{F}_S(\mathcal{H}_1)$ are the same as those satisfied by $\{\hat{\Omega}_2(\psi, \cdot)\} \in \mathcal{F}_S(\mathcal{H}_2)$ [12]. This will be the algebraic approach.

This formulation consists of taking the field, defined in Eq. (3.69), to an algebraic space, $\mathcal{A}(\mathbb{M})$, of observables of the Klein-Gordon field. An algebra $\mathcal{A}$ over the set of complex numbers $\mathbb{C}$ is defined as a complex vector space equipped with a bilinear and associative operation. Thus, $\mathcal{A}$ will be an algebra if [12, 27]

i. there is a map $\bullet$, so that $\bullet : \mathcal{A} \times \mathcal{A} \to \mathcal{A}$;

ii. to $(A_1, A_2, A_3) \in \mathcal{A}$, the associative operation $(A_1A_2)A_3 = A_1(A_2A_3)$ is satisfied.

Also, we say that $\mathcal{A}$ is a normed algebra if it has a well defined norm, such that if $a \in \mathcal{A}$, then

$$a \in \mathcal{A} \to ||a|| \in \mathbb{R}, \tag{3.77}$$

where $\mathbb{R}$ is the set of real numbers. Therefore, to any $a, b \in \mathcal{A}$ and $\gamma \in \mathbb{C}$, the norm must satisfy the following relations:

i. $||a|| \geq 0$;

ii. $||a|| = 0$ if, and only if, $a = 0$;

iii. $||\gamma a|| = |\gamma| \, ||a||$, where $|\gamma|$ is the module of the complex number γ;

iv. $||a + b|| \leq ||a|| + ||b||$;

v. $||ab|| \leq ||a|| \, ||b||$.

Moreover, if all Cauchy sequences on $\mathcal{A}$ converges, i.e., the elements of $\mathcal{A}$ are closer to each other than any given positive distance, the algebra $\mathcal{A}$ is said complete. A complete normed algebra is called a Banach algebra [12].

If $\mathcal{A}$ is a Banach algebra, an involution on it, i.e., a map which take us to the starting point when it is applied twice, is defined as

$$^*: \mathcal{A} \to \mathcal{A}, \tag{3.78}$$

such that, for any $a, b \in \mathcal{A}$ and $\gamma \in \mathbb{C}$, it must satisfy:

i. $(a+b)^* = a^* + b^*$;

ii. $(\gamma a)^* = \overline{\gamma} a^*$;

iii. $(ab)^* = b^*\, a^*$;

iv. $(a^*)^* = a$;

v. $||a^*|| = ||a||$.

A Banach algebra equipped with an involution is a **-algebra*. Lastly, if it also fulfill the property

vi. $||a^* a|| = ||a||^2, \forall a \in \mathcal{A}$,

we call it a *C^*-algebra* [12, 27].

This formulation is quite different from what we did before. Now, instead of defining operators which act upon vectors in a Hilbert space, we will build operators as elements of $\mathcal{A}(\mathbb{M})$ over which the states will act by identifying a number to each one of it [12]. In this way, we can proceed with the quantization procedure writing the field operator defined in Eq. (3.69) as an element of $\mathcal{A}$, such that, to any $f \in C_0^\infty(\mathbb{M})$, we will have a map [11, 12]

$$\phi : f \in C_0^\infty(\mathbb{M}) \to \hat{\phi}(f) \in \mathcal{A}(\mathbb{M}), \tag{3.79}$$

where $\mathcal{A}(\mathbb{M})$ is a *-algebra. Moreover, the operator $\hat{\phi}(f)$ satisfies the following properties [11, 13]:

1. $\hat{\phi}^*(f) = \hat{\phi}(f)$, for any $f \in \mathcal{T}$;
2. $\hat{\phi}([\nabla^\mu \nabla_\mu - m^2]\, g) = 0$, for some $f = [\nabla^\mu \nabla_\mu - m^2]\, g$ and $g \in \mathcal{T}$;
3. $\left[\hat{\phi}(f), \hat{\phi}(g)\right] = -i\Delta(f,g)\mathbb{I}$.

Therefore, as we said before, a quantum state in this formulation will act upon an operator, which is defined in an algebraic space, by associating a number to each one of it. So we must define a quantum state, ω, as a linear functional on an algebraic space, $\mathcal{A}(\mathbb{M})$, such that $\omega : \mathcal{A}(\mathbb{M}) \to \mathbb{C}$. As a quantum state, it must satisfy the positivity condition, i.e., $\omega(a^* a) \geq 0 \, \forall \, a \in \mathcal{A}(\mathbb{M})$, and the normalization condition, i.e., $\omega(\mathbb{I}) = 1$ where $\mathbb{I}$ is the identity element of $\mathcal{A}(\mathbb{M})$ [11, 12].

In order to facilitate our formulation of QFT in curved spacetimes, we will use an alternative choice for the *-algebra of fundamental observables, $\mathcal{W}(\mathbb{M})$, which is called Weyl algebra. Therefore, in this construction, we define fundamental observables as the operators [12]

$$\hat{W}(Ef) \equiv e^{i\hat{\phi}(f)}, \tag{3.80}$$

which satisfies the Weyl relations,

$$\hat{W}^*(Ef) = \hat{W}(-Ef), \tag{3.81}$$

$$\hat{W}\left[E(\nabla^\mu \nabla_\mu - m^2)f\right] = \mathbb{I}, \tag{3.82}$$

$$\hat{W}(Ef)\hat{W}(Eg) = e^{i\Delta(f,g)/2}\,\hat{W}(Ef + Eg), \tag{3.83}$$

for all $f, g \in \mathcal{T}$.

If we follow the same reasoning we did when we defined the quantum states before, we can say that the states which will arise from this quantum field construction must be a positive and normalized linear functional ω, such that $\omega : \mathcal{W}(\mathbb{M}) \to \mathbb{C}$, since $\mathcal{W}(\mathbb{M})$ is also a *-algebra.

Lastly, let $(\mathcal{S}, \Omega)$ be a vector space of real solutions to Eq. (3.36), with well posed initial values in Σ_t. The specification of a Hilbert space in the space of complex solutions, $\mathcal{S}^{\mathbb{C}}$ allow us to define an inner product, μ, on the subspace of real solutions, $\mathcal{S}$, such that $\mu : \mathcal{S} \times \mathcal{S} \to \mathbb{R}$. The inner product between two solutions, ψ_1 and ψ_2, with $\psi_1, \psi_2 \in \mathcal{S}$, will be [12]

$$\mu(\psi_1, \psi_2) = \text{Re}\,\langle K\psi_1 | K\psi_2 \rangle_{\text{KG}} = \text{Im}\,\Omega(\overline{K\psi_1}, K\psi_2), \tag{3.84}$$

where we used Eq. (3.48) and $K : \mathcal{S} \to \mathcal{H}$. Also, the imaginary part of the Klein-Gordon inner product can be written as

$$\begin{aligned} \text{Im}\,\langle K\psi_1 | K\psi_2 \rangle_{\text{KG}} &= -\text{Re}\,\Omega(\overline{K\psi_1}, K\psi_2) \\ &= -\frac{1}{2}\Omega(\overline{K\psi_1}, K\psi_2) - \frac{1}{2}\Omega(K\psi_1, \overline{K\psi_2}) \\ &= -\frac{1}{2}\Omega(\psi_1, \psi_2), \end{aligned} \tag{3.85}$$

where $\psi_1 = K\psi_1 + \overline{K\psi_1}$ and $\psi_2 = K\psi_2 + \overline{K\psi_2}$ for all $\psi_1, \psi_2 \in \mathcal{S}$ [12]. Then, if we use that

$$\langle K\psi_1 | K\psi_2 \rangle_{\text{KG}} = \text{Re}\;\langle K\psi_1 | K\psi_2 \rangle_{\text{KG}} + \text{Im}\;\langle K\psi_1 | K\psi_2 \rangle_{\text{KG}}\,, \tag{3.86}$$

along with Eqs. (3.84) and (3.85), we get

$$\langle K\psi_1|K\psi_2\rangle_{\text{KG}} = \mu(\psi_1,\psi_2) - \frac{1}{2}\Omega(\psi_1,\psi_2). \tag{3.87}$$

Now, we can utilize Schwarz inequality,

$$||x_1||^2||x_2||^2 \geq |\,\langle x_1|x_2\rangle\,|^2 \geq |\text{Im}\,\langle x_1|x_2\rangle\,|^2, \tag{3.88}$$

with $x_1 = K\psi_1$ and $x_2 = K\psi_2$, to write

$$||K\psi_1||^2||K\psi_2||^2 \geq |\,\langle K\psi_1|K\psi_2\rangle\,|^2 \geq |\text{Im}\,\langle K\psi_1|K\psi_2\rangle_{\text{KG}}\,|^2, \tag{3.89}$$

and

$$\mu(\psi_1,\psi_1)\mu(\psi_2,\psi_2) \geq \frac{1}{4}|\Omega(\psi_1,\psi_2)|^2, \tag{3.90}$$

so that, μ is positive-definite.

Thus, considering the first property of E, we can make $\psi_1 = Ef$ and $\psi_2 = Eg$, to any $f, g \in \mathcal{T}$, such that, the above relation becomes

$$\mu(Ef,Ef)\mu(Eg,Eg) \geq \frac{1}{4}|\Omega(Ef,Eg)|^2. \tag{3.91}$$

To a real inner product, μ, on $\mathcal{S}$, which satisfies the relation (3.91), we can define an algebraic state, ω_μ, associated with it, by

$$\omega_\mu\left[\hat{W}(Ef)\right] \equiv e^{-\mu(Ef,Ef)/2}. \tag{3.92}$$

Also, if μ satisfies (3.91), we call ω a quasi-free state [12].

Therefore, the fundamental observables of the quantum field theory in curved spacetimes will be defined as elements of the Weyl algebra. This formulation is useful because it will not demand us to take any preferred choice, in the sense that, even to unitarily inequivalent Hilbert space constructions, generated by preferred choices of μ, which satisfies Eq. (3.91), the algebraic spaces which they give rise, are isomorphic.

In the next chapter we will apply this algebraic formulation to a system composed by a scalar field and two qubits, one for Alice and another one for Bob, observers at arbitrary points in the spacetime. Our goal is to calculate the classical capacity of the quantum channel and the energetic balance of the communication process in a general globally hyperbolic spacetime.

Chapter 4

Energy Cost for the Transmission of Information for a Relativistic Quantum Communication Channel

In the previous chapters we investigated some of the main concepts of information theory, such as information measures, data compression and also channel capacity. Then, based on the ideas of general relativity, we described a quantum field theory in globally hyperbolic spacetimes. Now, we are ready to put all of these concepts together and describe mathematically the quantum channel used by Alice and Bob. Additionally, we will calculate the capacity of the quantum channel as well as the energy cost for transmitting information on it.

First of all, we will give a mathematical description of the communication channel between Alice and Bob. This will be made by considering a system composed of two-qubits and a quantum scalar field. We will trace out the field degrees of freedom of the initial system state and then take the partial trace over the Alice's qubit A. By doing this, we obtain the final state of Bob's qubit, which can be defined as a map that describes the communication channel. After doing this, we will calculate the channel capacity, i.e., the maximum amount of information which can be reliably transmitted per use of the channel.

We will dedicate the second half of the chapter to describe the quantum field theory restricted to null sub-manifolds. It will be done because, as we will see, we will assume a massless field and, since we are considering the relativistic context, when we evaluate the total Hamiltonian from the infinity null past to the infinity null future — these concepts will be discussed later — we won't need to worry about the interaction Hamiltonian, once the interactions happens for a limited amount of time. In this way, we will calculate the total system energy variation and investigate each one of its contributions.

4.1 The Quantum Communication Channel

Before we look for a mathematical description of the communication channel, let us first review the communication procedure. Thus, consider that Alice and Bob want to communicate with each other and they will use a quantum field ϕ as a noisy communication channel, i.e., the information Bob receives is not completely reliable.

As a means to imprint some information on the field state, Alice needs to interact her qubit with the quantum field for a certain amount of time, the same must be done by Bob in order to get the information. Therefore, consider that Alice's qubit, which is associated with the Hilbert space $\mathcal{H}_A$, is prepared in the state $\rho_{-\infty}^A$ and interacts with the quantum field, that is in some quasi-free state ω_μ, for a limited amount of time Δt_A, relative to the Cauchy surface Σ_t.

After Alice switches off her qubit interaction with the field, Bob will switch on his qubit interaction. It is initially prepared in the state $\rho_{-\infty}^B$, which is associated with the Hilbert space $\mathcal{H}_B$. As Alice's, Bob's qubit will interact with the field for a limited amount of time Δt_B [11]. The interaction can't be done for too long, in order to avoid information losses due to decoherence. Additionally, Bob knows the field state even before Alice interacts her qubit with it, so, after performing the measurement, he will be able to tell how the field state has changed after Alice's interaction.

Therefore, we expect that the initial state of the qubits is given by $\rho_{-\infty}^A \otimes \rho_{-\infty}^B$, such that, the initial state of the system, which is composed by the two qubits and the field, will be given by $\rho_{-\infty} \equiv \rho_{-\infty}^A \otimes \rho_{-\infty}^B \otimes \rho_\omega$, where ρ_ω is the density operator associated with the algebraic state ω_μ, such that $\omega_\mu(\hat{W}(Ef)) \equiv \mathrm{Tr}\left(\rho_\omega \hat{W}(Ef)\right)$. Thus, the total Hamiltonian of the system will be given by the Hamiltonian of the field plus the one associated with the interaction between the qubits and the field

$$H(t) \equiv H_\phi(t) + H_{\text{int}}(t), \tag{4.1}$$

where $H_\phi(t)$ is the field Hamiltonian, which is defined in Eq. (3.41), and $H_{\text{int}}(t)$ is the interaction Hamiltonian [11, 13]

$$H_{\text{int}}(t) \equiv \sum_j \epsilon_j(t) \int_{\Sigma_t} d^3\mathbf{x}\sqrt{-g}\, \psi_j(t,\mathbf{x})\phi(t,\mathbf{x}) \otimes \sigma_j^z. \tag{4.2}$$

$j = A, B$ labels Alice and Bob's qubits, respectively, while $\epsilon_j(t) \in C_0^\infty(\mathbb{R})$ is a real and compact support coupling function that sustain the interaction of the qubit j with the field for a finite amount of time Δt_j [11]. Also, $\psi_j(t,\mathbf{x})$ is a real and smooth function, which satisfies $\psi_j|_{\Sigma_t} \in C_0^\infty(\Sigma_t)$, and limits the interaction of each qubit to the vicinity of its own world line [11]. Lastly, σ_j^z is the Pauli matrix on z-dimension associated with the qubit j, which appears because of the field's interaction with the two-level systems.

Now that we have the initial state of the system, as well as its Hamiltonian, our next job is to evolve it in time, relative to the Cauchy surface Σ_t, so we can find the final system state. Thus, in the interaction representation, we can define the time-evolution operator associated with the interaction Hamiltonian as [13]

$$U = T\exp\left\{-i\int_{-\infty}^{\infty} H_{\text{int}}(t)dt\right\}, \tag{4.3}$$

where T is the time ordering operator relative to t. This operator is responsible to take any product of operators, defined in different times, and sort them in descending order in time. It is important because we have a time order of interaction with the field.

The derivative of U, relative to t, can be written as

$$U' = T\frac{\mathrm{d}}{\mathrm{d}t}\exp\left\{-i\int_{-\infty}^{\infty} H_{\text{int}}(t)dt\right\} \tag{4.4}$$

$$= -iH_{\text{int}}(t)U. \tag{4.5}$$

The Magnus theorem says that, the solution to the linear evolution equation

$$X'(t) = A(t)X(t), \tag{4.6}$$

with a well posed initial value, can be expressed as an exponential of a function $\Omega(t)$

$$X(t) = e^{\Omega(t)}. \tag{4.7}$$

If we identify $A(t) = -iH_{\text{int}}(t)$ such that $X(t) = U$, we can write

$$U = e^{\Omega(t)}, \tag{4.8}$$

where $\Omega(t)$ is defined by the expansion

$$\Omega(t) \equiv \sum_{n=1}^{\infty} \Omega_n, \tag{4.9}$$

with Ω_n being an operator of order n in $H_{\text{int}}(t)$ [11].

We can calculate the expansion of $\Omega(t)$ by

$$\Omega(t) = \int_{t_0}^{t_1} A(t)dt - \frac{1}{2}\int_{t_0}^{t_1}\left[\int_{t_0}^{t} A(t')dt', A(t)\right] dt + ... \tag{4.10}$$

Only the first two terms will be relevant to us, since the terms with $n \geq 0$ will depends on the commutator between $\hat{\phi}(f_j)$ and $\mathbb{I}$, which is zero. Thus, for t going from $-\infty$ to ∞, the first term of the sum will be

$$\Omega_1 = -i\int_{-\infty}^{\infty} H_{\text{int}}(t)dt. \tag{4.11}$$

To the second term can be written as

$$\int_{t_0}^{t_1}\left[\int_{t_0}^{t} A(t')dt', A(t)\right] dt = \int_{-\infty}^{\infty} dt \int_{-\infty}^{t} dt' \left[H_{\text{int}}(t), H_{\text{int}}(t')\right]. \tag{4.12}$$

Therefore,

$$\Omega_2 = -\frac{1}{2}\int_{-\infty}^{\infty} dt \int_{-\infty}^{t} dt' \left[H_{\text{int}}(t), H_{\text{int}}(t')\right]. \tag{4.13}$$

If we substitute Eq. (4.2) into Ω_1, we get

$$\begin{aligned}
\Omega_1 &= -i\int_{-\infty}^{\infty} dt \left[\sum_j \epsilon_j(t)\int_{\Sigma_t} d^3\mathbf{x}\sqrt{-g}\,\psi_j(t,\mathbf{x})\phi(t,\mathbf{x})\otimes\sigma_j^z\right] && (4.14)\\
&= -i\sum_j \int_{\mathbb{M}} d^4\mathbf{x}\epsilon_j(t)\sqrt{-g}\,\psi_j(\mathbf{x})\phi(\mathbf{x})\otimes\sigma_j^z && (4.15)\\
&= -i\sum_j \int_{\mathbb{M}} d^4\mathbf{x}\sqrt{-g}\,f_j(\mathbf{x})\phi(\mathbf{x})\otimes\sigma_j^z, && (4.16)
\end{aligned}$$

where we defined

$$f_j(\mathbf{x}) \equiv \epsilon_j(t)\psi_j(\mathbf{x}). \tag{4.17}$$

Also, from the previous chapter, we can use Eq. (3.57) along with Eq. (3.69) to write

$$\hat{\phi}(f_j) \equiv \int_{\mathbb{M}} d^4\mathbf{x}\sqrt{-g}\, f_j(\mathbf{x})\phi(\mathbf{x}), \tag{4.18}$$

such that the first term of the expansion of $\Omega(t)$ becomes

$$\Omega_1 = -i\sum_j \hat{\phi}(f_j) \otimes \sigma_j^z, \tag{4.19}$$

for $j = A, B$.

Again, if we substitute Eq. (4.2) into Ω_2, we obtain [11]

$$\begin{aligned}\Omega_2 = -\frac{1}{2}\int_{-\infty}^{\infty} dt \int_{-\infty}^{t} dt' \Bigg[\sum_j \epsilon_j(t)\int_{\Sigma_t} d^3\mathbf{x}\sqrt{-g}\,\psi_j(t,\mathbf{x})\phi(t,\mathbf{x}) \otimes \sigma_j^z,\\ \sum_i \epsilon_i(t')\int_{\Sigma_{t'}} d^3\mathbf{x'}\sqrt{-g'}\,\psi_j(t',\mathbf{x'})\phi(t',\mathbf{x'}) \otimes \sigma_i^z\Bigg],\end{aligned} \tag{4.20}$$

to $i = A, B$. Defining

$$\Xi \equiv \frac{1}{2}\sum_j \int_{-\infty}^{\infty} dt\epsilon_j(t)\int_{-\infty}^{t} dt'\epsilon_j(t')\Delta_j(t,t'), \tag{4.21}$$

with

$$\Delta_j(t,t') \equiv \int_{\Sigma_t} d^3\mathbf{x}\sqrt{-g}\int_{\Sigma_{t'}} d^3\mathbf{x'}\sqrt{-g'}\,\psi_j(t,\mathbf{x})\Delta(x,x')\psi_j(t',\mathbf{x'}) \tag{4.22}$$

and

$$\left[\hat{\phi}(x), \hat{\phi}(x')\right] \equiv -i\Delta(x,x')\mathbb{I}, \tag{4.23}$$

we can use it and open the commutator in Eq. (4.20), to get [11]

$$\Omega_2 = i\Xi\mathbb{I} - \frac{i}{2}\Delta(f_A,f_B)\sigma_A^z \otimes \sigma_B^z, \tag{4.24}$$

where, on the first term we consider the case $j = i$ and the second one is when $j \neq i$.

Now we can substitute Eqs. (4.19) and (4.24) into Eq. (4.8), by using the expansion defined in Eq. (4.9), to get the time evolution operator

$$U = e^{-i\sum_j \hat{\phi}(f_i)\otimes\sigma_j^z}\, e^{i\Xi\mathbb{I} - \frac{i}{2}\Delta(f_A,f_B)\sigma_A^z\otimes\sigma_B^z} \tag{4.25}$$

or

$$U = e^{-i\hat{\phi}(f_A)\otimes\sigma_A^z - i\hat{\phi}(f_B)\otimes\sigma_B^z} e^{i\Xi\mathbb{I} - \frac{i}{2}\Delta(f_A,f_B)\sigma_A^z\otimes\sigma_B^z}. \tag{4.26}$$

If we use the Baker–Campbell–Hausdorff (BCH) formula

$$e^{x+y} = e^x e^y e^{-\frac{1}{2}[x,y]}, \tag{4.27}$$

which is valid when $[x,[x,y]] = [y,[x,y]] = 0$, and replacing x and y as

$$x = -i\hat{\phi}(f_A) \otimes \sigma_A^z, \tag{4.28}$$

$$y = -i\hat{\phi}(f_B) \otimes \sigma_B^z, \tag{4.29}$$

the commutator between these operators will be

$$\begin{aligned} [x,y] &= \left[-i\hat{\phi}(f_A) \otimes \sigma_A^z, -i\hat{\phi}(f_B) \otimes \sigma_B^z\right] \\ &= -\left[\hat{\phi}(f_A), \hat{\phi}(f_B)\right] \sigma_A^z \otimes \sigma_B^z \\ &= i\Delta(f_A, f_B)\mathbb{I}\sigma_A^z \otimes \sigma_B^z, \end{aligned} \tag{4.30}$$

and the BCH formula,

$$e^{-i\hat{\phi}(f_A)\otimes\sigma_A^z - i\hat{\phi}(f_B)\otimes\sigma_B^z} = e^{-i\hat{\phi}(f_A)\otimes\sigma_A^z} e^{-i\hat{\phi}(f_B)\otimes\sigma_B^z} e^{-\frac{i}{2}\Delta(f_A,f_B)\sigma_A^z\otimes\sigma_B^z}. \tag{4.31}$$

It can be used in Eq. (4.26) to write it as

$$U = e^{i\Xi\mathbb{I}} e^{-i\hat{\phi}(f_A)\otimes\sigma_A^z} e^{-i\hat{\phi}(f_B)\otimes\sigma_B^z} e^{-i\Delta(f_A,f_B)\sigma_A^z\otimes\sigma_B^z}, \tag{4.32}$$

where we explored that $[\hat{\phi}(f_j) \otimes \sigma_j^z, \mathbb{I}] = 0$.

Now that we have the time evolution operator, we can use it to evolve the initial state of the system $\rho_{-\infty}$ and obtain the final state of the two-qubits plus field $\rho_{+\infty} = U\rho_{-\infty}U^\dagger$. Later, we will use this state to compute the total energy variation of the system, but for now, we are interested in describing the communication channel used by Alice and Bob.

The expression which describes the communication channel should relate the initial state of Alice's qubit to the final state of Bob's qubit, i.e., it is a map that take us from the state $\rho_{-\infty}^A$ to the state ρ^B. In order to find the final state of Bob's qubit, we need to trace out the field and Alice's qubit degrees of freedom from the system state after the communication process has ended. In this way, we can define Bob's qubit final state as

$$\rho^B \equiv \mathrm{Tr}_{A,\phi}(U\rho_{-\infty}U^\dagger) \equiv \mathrm{Tr}_{A,\phi}(U\rho_{-\infty}^A \otimes \rho_{-\infty}^B \otimes \rho_\omega U^\dagger). \tag{4.33}$$

Before we proceed with this calculation, let us first make the following expansion

$$\exp\Big\{-i\hat{\phi}(f_j)\otimes\sigma_j^z\Big\} = 1 - i\hat{\phi}(f_j)\otimes\sigma_j^z - \frac{1}{2!}[\hat{\phi}(f_j)\otimes\sigma_j^z]^2 + \frac{i}{3!}[\hat{\phi}(f_j)\otimes\sigma_j^z]^3 - \dots, \tag{4.34}$$

where we used the property

$$(\sigma_j^z)^n = \begin{cases} 1 & \text{if } n \text{ is even;} \\ \sigma_j^z & \text{if } n \text{ is odd.} \end{cases} \tag{4.35}$$

Thus, if we write this exponential function in terms of sines and cosines, we get

$$\exp\Big\{-i\hat{\phi}(f_j)\otimes\sigma_j^z\Big\} = \cos\hat{\phi}(f_j)\mathbb{I} - i\sin[\hat{\phi}(f_j)]\otimes\sigma_j^z, \tag{4.36}$$

where the operators $\cos\hat{\phi}(f_j)$ and $\sin\hat{\phi}(f_j)$ are defined according to

$$\cos\hat{\phi}(f_j) \equiv \frac{1}{2}\left[e^{i\hat{\phi}(f_j)} + e^{-i\hat{\phi}(f_j)}\right] \tag{4.37}$$

and

$$\sin\hat{\phi}(f_j) \equiv \frac{1}{2i}\left[e^{i\hat{\phi}(f_j)} - e^{-i\hat{\phi}(f_j)}\right]. \tag{4.38}$$

We can use the Weyl operator in Eq. (3.80) to rewrite the above equations as

$$\cos\hat{\phi}(f_j) \equiv \frac{1}{2}\left[\hat{W}(Ef_j) + \hat{W}(-Ef_j)\right] \tag{4.39}$$

and

$$\sin\hat{\phi}(f_j) \equiv \frac{1}{2i}\left[\hat{W}(Ef_j) - \hat{W}(-Ef_j)\right]. \tag{4.40}$$

Now, using the above definitions, we can trace out the field degrees of freedom of the system state after the communication process has ended. This process will give us the state, ρ^{AB}, without the field degrees of freedom. After this process is done we will need to trace out Alice's qubit degrees of freedom, as Eq. (4.33) suggests, in order to find Bob's qubit final state.

Therefore, the state ρ^{AB} will be [11, 13]

$$\rho^{AB} \equiv \mathrm{Tr}_\phi\left(U\rho_{-\infty}^A\otimes\rho_{-\infty}^B\otimes\rho_\omega U^\dagger\right) \tag{4.41}$$

$$\begin{aligned}= \mathrm{Tr}_\phi\Big\{&e^{i\Xi\mathbb{I}}e^{-i\hat{\phi}(f_A)\otimes\sigma_A^z}e^{-i\hat{\phi}(f_B)\otimes\sigma_B^z}e^{-i\Delta(f_A,f_B)\sigma_A^z\otimes\sigma_B^z}\rho_{-\infty}^A\otimes\rho_{-\infty}^B\otimes\rho_\omega\times \\ &\times e^{-i\Xi\mathbb{I}}e^{i\Delta(f_A,f_B)\sigma_A^z\otimes\sigma_B^z}e^{i\hat{\phi}(f_B)\otimes\sigma_B^z}e^{i\hat{\phi}(f_A)\otimes\sigma_A^z}\Big\}\end{aligned} \tag{4.42}$$

$$
\begin{aligned}
&= \mathrm{Tr}_\phi \Big\{ e^{-i\hat{\phi}(f_A)\otimes\sigma_A^z} e^{-i\hat{\phi}(f_B)\otimes\sigma_B^z} e^{-i\Delta(f_A,f_B)\sigma_A^z\otimes\sigma_B^z} (\rho_{-\infty}^A \otimes \rho_{-\infty}^B)\times \\
&\times e^{i\Delta(f_A,f_B)\sigma_A^z\otimes\sigma_B^z} \rho_\omega e^{i\hat{\phi}(f_B)\otimes\sigma_B^z} e^{i\hat{\phi}(f_A)\otimes\sigma_A^z} \Big\},
\end{aligned} \tag{4.43}
$$

from which we can define

$$
\tilde{\rho}^{AB} \equiv e^{-i\Delta(f_A,f_B)\sigma_A^z\otimes\sigma_B^z} (\rho_{-\infty}^A \otimes \rho_{-\infty}^B) e^{i\Delta(f_A,f_B)\sigma_A^z\otimes\sigma_B^z}, \tag{4.44}
$$

to write

$$
\rho^{AB} = \mathrm{Tr}_\phi \Big\{ e^{-i\hat{\phi}(f_A)\otimes\sigma_A^z} e^{-i\hat{\phi}(f_B)\otimes\sigma_B^z} \tilde{\rho}^{AB} \rho_\omega e^{i\hat{\phi}(f_B)\otimes\sigma_B^z} e^{i\hat{\phi}(f_A)\otimes\sigma_A^z} \Big\} \tag{4.45}
$$

$$
\begin{aligned}
&= \mathrm{Tr}_\phi \Big\{ \Big[\cos\hat{\phi}(f_A) - i\sin\hat{\phi}(f_A)\otimes\sigma_A^z\Big] \Big[\cos\hat{\phi}(f_B) - \\
&- \ i\sin\hat{\phi}(f_B)\otimes\sigma_B^z\Big] \tilde{\rho}^{AB}\rho_\omega \Big[\cos\hat{\phi}(f_B) + i\sin\hat{\phi}(f_B)\otimes\sigma_B^z\Big] \times \\
&\times \Big[\cos\hat{\phi}(f_A) + i\sin\hat{\phi}(f_A)\otimes\sigma_A^z\Big]\Big\}.
\end{aligned} \tag{4.46}
$$

Note that, by performing the multiplication, the first term will be

$$
\mathrm{Tr}_\phi \Big\{ \cos\hat{\phi}(f_A)\cos\hat{\phi}(f_B)\tilde{\rho}^{AB}\rho_\omega \cos\hat{\phi}(f_B)\cos\hat{\phi}(f_A) \Big\}, \tag{4.47}
$$

but, since we are tracing out only the field degrees of freedom, it becomes

$$
\mathrm{Tr}_\phi \Big\{ \cos\hat{\phi}(f_B)\cos\hat{\phi}(f_A)\cos\hat{\phi}(f_A)\cos\hat{\phi}(f_B)\rho_\omega \Big\} \tilde{\rho}^{AB}, \tag{4.48}
$$

where we used the cyclic property of trace. The calculation of the other terms will be similar. Now, applying the relation $\omega_\mu(\hat{W}(Ef)) \equiv \mathrm{Tr}\Big(\rho_\omega \hat{W}(Ef)\Big)$, as well as $\omega_\mu[\hat{W}(Ef)] \in \mathbb{R}_+$, and defining

$$
\Gamma_{\alpha\beta\gamma\delta} \equiv \omega_\mu(\mathcal{F}_\alpha[\phi(f_B)]\mathcal{F}_\beta[\phi(f_A)]\mathcal{F}_\gamma[\phi(f_A)]\mathcal{F}_\delta[\phi(f_B)]), \tag{4.49}
$$

where $\alpha,\beta,\gamma,\delta \in \{s,c\}$, $\mathcal{F}_s[\phi] \equiv \sin\phi$ and $\mathcal{F}_c[\phi] \equiv \cos\phi$, we will get, after some algebra,

$$
\begin{aligned}
\rho^{AB} &= \Gamma_{cccc}\tilde{\rho}^{AB} + \Gamma_{ssss}\sigma_A^z\otimes\sigma_B^z\tilde{\rho}^{AB}\sigma_A^z\otimes\sigma_B^z + \Gamma_{cssc}\sigma_A^z\tilde{\rho}^{AB}\sigma_A^z + \\
&+ \Gamma_{sccs}\sigma_B^z\tilde{\rho}^{AB}\sigma_B^z - \Gamma_{ccss}\sigma_A^z\otimes\sigma_B^z\tilde{\rho}^{AB} - \Gamma_{sscc}\tilde{\rho}^{AB}\sigma_A^z\otimes\sigma_B^z + \\
&+ \Gamma_{scsc}\sigma_A^z\tilde{\rho}^{AB}\sigma_B^z + \Gamma_{cscs}\sigma_B^z\tilde{\rho}^{AB}\sigma_A^z,
\end{aligned} \tag{4.50}
$$

so that, we can abbreviate it as [11]

$$
\begin{aligned}
\rho^{AB} &= \Gamma_{cccc}\tilde{\rho}^{AB} + \Gamma_{ssss}\sigma_A^z\otimes\sigma_B^z\tilde{\rho}^{AB}\sigma_A^z\otimes\sigma_B^z + \Gamma_{cssc}\sigma_A^z\tilde{\rho}^{AB}\sigma_A^z + \\
&+ \Gamma_{sccs}\sigma_B^z\tilde{\rho}^{AB}\sigma_B^z - \left(\Gamma_{sscc}\tilde{\rho}^{AB}\sigma_A^z\otimes\sigma_B^z + \mathrm{H.c.}\right) + \\
&+ \left(\Gamma_{cscs}\sigma_B^z\tilde{\rho}^{AB}\sigma_A^z + \mathrm{H.c.}\right),
\end{aligned} \tag{4.51}
$$

where H.c. stands for Hermitian conjugate. The $\Gamma_{\alpha\beta\gamma\delta}$ can be written by using the algebraic states defined by

$$\nu_j \equiv \omega_\mu(\hat{W}[E(2f_j)]) = e^{-2||Ef_j||^2}, \tag{4.52}$$

$$\nu_{AB}^+ \equiv \omega_\mu(\hat{W}[E(2f_A + 2f_B)]) = e^{-2||E(f_A+f_B)||^2} \tag{4.53}$$

and

$$\nu_{AB}^- \equiv \omega_\mu(\hat{W}[E(2f_A - 2f_B)]) = e^{-2||E(f_A-f_B)||^2}. \tag{4.54}$$

In order to do that, we need to use the Weyl relations, which were shown in the previous chapter, as well as Eqs. (4.37) and (4.38). We can calculate one term here and the others will follow the same procedure. Thus, if we substitute Eq. (4.38) into Γ_{ssss}, we get

$$\begin{aligned}
\Gamma_{ssss} &= \omega_\mu\left(\mathcal{F}_s[\phi(f_B)]\mathcal{F}_s[\phi(f_A)]\mathcal{F}_s[\phi(f_A)]\mathcal{F}_s[\phi(f_B)]\right) \\
&= \omega_\mu\left[\sin\phi(f_B)\sin\phi(f_A)\sin\phi(f_A)\sin\phi(f_B)\right] \\
&= \omega_\mu\Bigg(\frac{1}{2i}[\hat{W}(Ef_B) - \hat{W}(-Ef_B)]\,\frac{1}{2i}[\hat{W}(Ef_A) - \hat{W}(-Ef_A)]\times \\
&\times\frac{1}{2i}[\hat{W}(Ef_A) - \hat{W}(-Ef_A)]\,\frac{1}{2i}[\hat{W}(Ef_B) - \hat{W}(-Ef_B)]\Bigg) \\
&= \omega_\mu\Bigg(\frac{1}{16}\Big[\hat{W}(Ef_B)\hat{W}(Ef_A) - \hat{W}(Ef_B)\hat{W}(-Ef_A) - \\
&-\ \hat{W}(-E_fB)\hat{W}(Ef_A) + \hat{W}(-Ef_B)\hat{W}(-Ef_A)\Big]\times \\
&\times\Big[\hat{W}(Ef_A) - \hat{W}(-Ef_A)\Big]\Big[\hat{W}(Ef_B) - \hat{W}(-Ef_B)\Big]\Bigg).
\end{aligned} \tag{4.55}$$

If we use the Weyl relation (3.83) and proceed with the calculations of all terms, we find that

$$\begin{aligned}
\Gamma_{ssss} = \omega_\mu\Bigg(\frac{1}{16}\Big\{-&\Big[\hat{W}(2Ef_A) + \hat{W}(-2Ef_A)\Big]\Big[e^{2i\Delta(f_A,f_B)} + \\
+\ & e^{-2i\Delta(f_A,f_B)}\Big] + \Big[\hat{W}(2Ef_A + 2Ef_B) + \hat{W}(-2Ef_A - 2Ef_B)\Big] + \\
+\ & \Big[\hat{W}(2Ef_A - 2Ef_B) + \hat{W}(-2Ef_A + 2Ef_B)\Big] + 4 - \\
-\ & 2\Big[\hat{W}(2Ef_B) + \hat{W}(-2Ef_B)\Big]\Big\}\Bigg),
\end{aligned} \tag{4.56}$$

where we can implement the definitions in Eqs. (4.52), (4.53) and (4.54), to write

$$\Gamma_{ssss} = \frac{1}{8}(\nu_{AB}^+ + \nu_{AB}^-) + \frac{1}{4}\left(1 - \nu_B - \nu_A\cos\left[2\Delta(f_A, f_B)\right]\right). \tag{4.57}$$

By following the same procedure, we can calculate all the other terms of $\Gamma_{\alpha\beta\gamma\delta}$. The results are [11]

$$\begin{aligned}
\Gamma_{cccc} &= \frac{1}{8}(\nu_{AB}^{+} + \nu_{AB}^{-}) + \frac{1}{4}\left(1 + \nu_B + \nu_A \cos\left[2\Delta(f_A, f_B)\right]\right), && (4.58)\\
\Gamma_{cssc} &= -\frac{1}{8}(\nu_{AB}^{+} + \nu_{AB}^{-}) + \frac{1}{4}\left(1 + \nu_B - \nu_A \cos\left[2\Delta(f_A, f_B)\right]\right), && (4.59)\\
\Gamma_{sccs} &= -\frac{1}{8}(\nu_{AB}^{+} + \nu_{AB}^{-}) + \frac{1}{4}\left(1 - \nu_B + \nu_A \cos\left[2\Delta(f_A, f_B)\right]\right), && (4.60)\\
\Gamma_{sscc} &= -\frac{1}{8}(\nu_{AB}^{+} - \nu_{AB}^{-}) + \frac{i}{4}\nu_A \sin\left[2\Delta(f_A, f_B)\right], && (4.61)\\
\Gamma_{cscs} &= -\frac{1}{8}(\nu_{AB}^{+} - \nu_{AB}^{-}) - \frac{i}{4}\nu_A \sin\left[2\Delta(f_A, f_B)\right], && (4.62)
\end{aligned}$$

with $\Gamma_{ccss} = \Gamma_{scsc} = 0$.

Now, in order to find ρ^B, we just need to trace out Alice's qubit degrees of freedom of the state ρ^{AB} —for more details see [31]—,

$$\rho^B \equiv \mathrm{Tr}_A(\rho^{AB}), \tag{4.63}$$

which procedure is similar to that of the field. By doing it, we find Bob's qubit final state to be

$$\begin{aligned}
\rho^B = &\frac{1}{2}\left(1 + \nu_B \cos[2\Delta(f_A, f_B)]\right)\rho_{-\infty}^B + \frac{1}{2}\left(1 - \nu_B \cos[2\Delta(f_A, f_B)]\right)\sigma_B^z \rho_{-\infty}^B \sigma_B^z + \\
&+ \frac{i}{2}\nu_B \sin[2\Delta(f_A, f_B)]\,\langle\sigma_A^z\rangle_{\rho_{-\infty}^A}\,[\rho_{-\infty}^B, \sigma_B^z],
\end{aligned} \tag{4.64}$$

where $\langle\sigma_A^z\rangle_{\rho_{-\infty}^A} \equiv \mathrm{Tr}\{\rho_{-\infty}^A \sigma_A^z\}$. Hence, Eq. (4.64) represents Bob's qubit state after the communication protocol has ended, i.e., the final system state without the degrees of freedom of the field and Alice's qubit. Now, in order to write the map which describes the communication channel, we must find one which is linear, trace preserving, completely positive and take us from the Alice's qubit initial state, $\rho_{-\infty}^A$, to Bob's qubit final state, ρ^B. The properties as trace preserving and completely positive will guarantee us that the result will be a density operator, so we can calculate the probabilities of the outcomes of any measurement performed upon the system. We define this map as $\mathcal{E}(\rho_{-\infty}^A)$.

But before we proceed with this calculation, first we need to set Bob's qubit initial state. Since σ_B^z commutes with the total Hamiltonian it is not productive to write $\rho_{-\infty}^B$ in terms of its eigenstates, because if we do that, Bob won't get any information transmitted by Alice. Thus, we can choose an initial state for Bob's qubit in such a way that it will maximize the signalling amplitude of the communication process [11, 13].

Attempting to understand how to maximize the signalling amplitude, let us write the measure probability associated to Bob. Let $\rho^A_{-\infty k}$, with $k = (+,-)$, be the state Alice encodes the message that will be transmitted. Then, Bob performs a projective measure by using the operators

$$F^B_+ \equiv |x_+\rangle_B \, {}_B\langle x_+|, \tag{4.65}$$

$$F^B_- \equiv |x_-\rangle_B \, {}_B\langle x_-|, \tag{4.66}$$

where $|x_+\rangle_B$ and $|x_-\rangle_B$ are eigenvectors associated with σ^x_B, such that, $\sigma^x_B |x_\pm\rangle = \pm |x_\pm\rangle$. The probability that Bob receives the outcome $l = \pm$ after the measure is given by $p(l|k) \equiv \mathrm{Tr}\big(F^B_l \rho^B_k\big)$, where ρ^B_k is the Bob's qubit state in Eq. (4.64).

Therefore, the probability will be

$$\begin{aligned}
\mathrm{Tr}\big(F^B_{l'}\rho^B_k\big) = \sum_l \langle x_l|x_{l'}\rangle \, {}_B\langle x_{l'}| \bigg\{ &\frac{1}{2}(1+\nu_B\cos[2\Delta(f_A,f_B)])\rho^B_{-\infty} + \\
+&\frac{1}{2}(1-\nu_B\cos[2\Delta(f_A,f_B)])\sigma^z_B\rho^B_{-\infty}\sigma^z_B + \\
+&\frac{i}{2}\nu_B\sin[2\Delta(f_A,f_B)]\,\langle\sigma^z_A\rangle_{\rho^A_{-\infty k}}\,[\rho^B_{-\infty},\sigma^z_B]\bigg\} |x_l\rangle_B \,,
\end{aligned} \tag{4.67}$$

and

$$\begin{aligned}
\mathrm{Tr}\Big(F^B_{x_{l'}}\rho^B_k\Big) = {}_B\langle x_+| \bigg\{ &\frac{1}{2}(1+\nu_B\cos[2\Delta(f_A,f_B)])\rho^B_{-\infty} + \\
+&\frac{1}{2}(1-\nu_B\cos[2\Delta(f_A,f_B)])\sigma^z_B\rho^B_{-\infty}\sigma^z_B + \\
+&\frac{i}{2}\nu_B\sin[2\Delta(f_A,f_B)]\,\langle\sigma^z_A\rangle_{\rho^A_{-\infty k}}\,[\rho^B_{-\infty},\sigma^z_B]\bigg\} |x_+\rangle_B + \\
+ {}_B\langle x_-| \bigg\{ &\frac{1}{2}(1+\nu_B\cos[2\Delta(f_A,f_B)])\rho^B_{-\infty} + \\
+&\frac{1}{2}(1-\nu_B\cos[2\Delta(f_A,f_B)])\sigma^z_B\rho^B_{-\infty}\sigma^z_B + \\
+&\frac{i}{2}\nu_B\sin[2\Delta(f_A,f_B)]\,\langle\sigma^z_A\rangle_{\rho^A_{-\infty k}}\,[\rho^B_{-\infty},\sigma^z_B]\bigg\} |x_-\rangle_B \,.
\end{aligned} \tag{4.68}$$

Using the definitions

$$|x_+\rangle_B \equiv \frac{1}{\sqrt{2}}(|+\rangle_B + |-\rangle_B), \tag{4.69}$$

and

$$|x_-\rangle_B \equiv \frac{1}{\sqrt{2}}(|+\rangle_B - |-\rangle_B), \tag{4.70}$$

where $\sigma^z |\pm\rangle_B \equiv \pm |\pm\rangle_B$. We can compute the above equation, so that the probability $p(l|k)$ is given by [11]

$$p(l|k) = \frac{1}{2}\left(1 + l\nu_B \Lambda_k\right), \tag{4.71}$$

where

$$\Lambda_k \equiv 2\,\mathrm{Re}\Big\{\beta_B \Big(\cos[2\Delta(f_A, f_B)] - i\,\langle\sigma_A^z\rangle_{\rho^A_{-\infty k}} \sin[2\Delta(f_A, f_B)]\Big)\Big\} \tag{4.72}$$

and $\beta_B = {}_B\langle +|\rho^B_{-\infty}|-\rangle_B$ [11].

By investigating Eq. (4.72) we see that we have two different contributions to the probability $p(l|k)$. Each one depends on the causal relation between the qubits, since they depend on $\Delta(f_A, f_B)$. If Bob is on Alice's causal future, then the communication is possible, such that $\Delta(f_A, f_B) \neq 0$, otherwise, we should have $\Delta(f_A, f_B) = 0$.

In order to represent the signalling, the contributions in Eq. (4.72) must depend on Alice's qubit initial state, since it is the state which the message will be encoded. Thus, we see that only the second one depends on it. So, if we want the communication to be as good as possible, this quantity must be maximized. Thus, Bob's qubit initial state, which will maximizes the signalling and, consequently, the channel capacity, is [11]

$$\rho^B_{-\infty} \equiv |y_+\rangle_B\,{}_B\langle y_+|, \tag{4.73}$$

with

$$|y_+\rangle \equiv \frac{1}{\sqrt{2}}(|+\rangle_B + i\,|-\rangle_B). \tag{4.74}$$

Since $|y_+\rangle_B$ is a superposition of the energy eigenstates it doesn't necessarily commutes with the Hamiltonian, so we can use it. Therefore

$$\begin{aligned}
\beta_B &= {}_B\langle +|\rho^B_{-\infty}|-\rangle_B \\
&= {}_B\langle +|\left[\frac{1}{2}(|+\rangle_B + i\,|-\rangle_B)({}_B\langle +| - i\,{}_B\langle -|)\right]|-\rangle_B \\
&= -\frac{i}{2}.
\end{aligned} \tag{4.75}$$

Substituting this result into Eq. (4.72) we obtain

$$\begin{aligned}\Lambda_k &= 2\,\mathrm{Re}\Big\{-\frac{i}{2}\left(\cos[2\Delta(f_A,f_B)] - i\,\langle\sigma_A^z\rangle_{\rho^A_{-\infty k}}\sin[2\Delta(f_A,f_B)]\right)\Big\}\\ &= 2\,\mathrm{Re}\Big\{-\frac{i}{2}\cos[2\Delta(f_A,f_B)] - \frac{1}{2}\langle\sigma_A^z\rangle_{\rho^A_{-\infty k}}\sin[2\Delta(f_A,f_B)]\Big\}\\ &= -\langle\sigma_A^z\rangle_{\rho^A_{-\infty k}}\sin[2\Delta(f_A,f_B)]. \end{aligned}\tag{4.76}$$

Then, the probability of the outcome l will be

$$p(l|k) = \frac{1}{2}\left(1 - l\nu_B\,\langle\sigma_A^z\rangle_{\rho^A_{-\infty k}}\sin[2\Delta(f_A,f_B)]\right). \tag{4.77}$$

Thus, Bob's qubit initial state which will maximize the signalling amplitude is the one defined in Eq. (4.73). Also, if we consider that Alice has encoded the message in the state $\rho^A_{-\infty}$, then, by using Eq. (4.77), the probability of error will be

$$p_e \equiv \frac{1}{2}\left(1 + \nu_B\,\langle\sigma_A^z\rangle_{\rho^A_{-\infty k}}\sin[2\Delta(f_A,f_B)]\right). \tag{4.78}$$

In this way, Bob's qubit initial state is defined and we can turn our attention back to the map which describes the quantum channel. In order to describe $\mathcal{E}(\rho^A_{-\infty})$ we can use the Kraus decomposition since it will take us from $\mathcal{H}_A$ to $\mathcal{H}_B$. The Kraus theorem ensures that any quantum operation which, in our case, is $\mathcal{E}(\rho^A_{-\infty})$, on a state $\rho^A_{-\infty}$ can always be written as

$$\mathcal{E}(\rho^A_{-\infty}) = \sum_{\mu=0}^{3} M_\mu \rho^A_{-\infty} M_\mu^\dagger, \tag{4.79}$$

for the operators M_μ which satisfies the completeness relation

$$\sum_{\mu=0}^{3} M_\mu^\dagger M_\mu = \mathbb{I}, \tag{4.80}$$

where $\mathbb{I}$ is the identity operator. We can define M_μ — for more details see [5] — according to [11]

$$M_0 \equiv \frac{1}{2}\sqrt{\frac{1-\nu_B^2}{p_e}}\,|+\rangle_{B\,A}\langle 0|, \tag{4.81}$$

$$M_1 \equiv \frac{1}{2}\sqrt{\frac{1-\nu_B^2}{1-P_e}}\,|+\rangle_{B\,A}\langle 1|, \tag{4.82}$$

$$M_2 \equiv \frac{i\nu_B}{2\sqrt{p_e}}\cos[2\Delta(f_A,f_B)]\,|+\rangle_{B\,A}\langle 0| + \sqrt{p_e}\,|-\rangle_{B\,A}\langle 0|, \tag{4.83}$$

$$M_3 \equiv \frac{i\nu_B}{2\sqrt{1-p_e}}\cos[2\Delta(f_A,f_B)]\,|+\rangle_{B\,A}\langle 1| + \sqrt{1-p_e}\,|-\rangle_{B\,A}\langle 1|. \tag{4.84}$$

Now, our final job is to substitute the above definitions into Eq. (4.79) and find the final form of $\mathcal{E}(\rho^A_{-\infty})$. By doing this we get

$$\begin{aligned}
\mathcal{E}(\rho^A_{-\infty}) = & \frac{1}{4}\left(\frac{1-\nu_B^2}{p_e}\right)|+\rangle_{B\,A}\langle 0|\rho^A_{-\infty}|0\rangle_{A\,B}\langle +|\sqrt{\frac{1-\nu_B^2}{p_e}}+ \\
& + \frac{1}{4}\left(\frac{1-\nu_B^2}{1-p_e}\right)|+\rangle_{B\,A}\langle 1|\rho^A_{-\infty}|1\rangle_{A\,B}\langle +|+ \\
& + \left[\frac{i\nu_B}{2\sqrt{p_e}}\cos[2\Delta(f_A,f_B)]\,|+\rangle_{B\,A}\langle 0| + \sqrt{p_e}\,|-\rangle_{B\,A}\langle 0|\right]\rho^A_{-\infty}\times \\
& \times\left[-\frac{i\nu_B}{2\sqrt{p_e}}\cos[2\Delta(f_A,f_B)]\,|0\rangle_{A\,B}\langle +| + \sqrt{p_e}\,|0\rangle_{A\,B}\langle -|\right]+ \\
& + \left[\frac{i\nu_B}{2\sqrt{1-p_e}}\cos[2\Delta(f_A,f_B)]\,|+\rangle_{B\,A}\langle 1| + \sqrt{1-p_e}\,|-\rangle_{B\,A}\langle 0|\right]\rho^A_{-\infty}\times \\
& \times\left[-\frac{i\nu_B}{2\sqrt{1-p_e}}\cos[2\Delta(f_A,f_B)]\,|1\rangle_{A\,B}\langle +| + \sqrt{1-p_e}\,|1\rangle_{A\,B}\langle -|\right].
\end{aligned} \tag{4.85}$$

Then, by performing the operations and simplifying some terms, we have

$$\begin{aligned}
\mathcal{E}(\rho^A_{-\infty}) = & \frac{1}{2}\left[(1-\nu_B\sin[2\Delta(f_A,f_B)])\,{}_A\langle 0|\rho^A_{\infty}|0\rangle_A+\right. \\
& \left.+(1+\nu_B\sin[2\Delta(f_A,f_B)])\,{}_A\langle 1|\rho^A_{-\infty}|1\rangle_A\right]|+\rangle_{B\,B}\langle +|+ \\
& + \frac{1}{2}\left[1+\nu_B\sin\sin\left[2\Delta(f_A,f_B)\right]\left({}_A\langle 0|\rho^A_{-\infty}|0\rangle_A - {}_A\langle 1|\rho^A_{-\infty}\right)|1\rangle_A\right]|-\rangle_{B\,B}\langle -| \\
& + \left(\frac{i\nu_B}{2}\cos[2\Delta(f_A,f_B)]\,|+\rangle_{B\,B}\langle -| + H.c\right)
\end{aligned} \tag{4.86}$$

$$\begin{aligned}
\mathcal{E}(\rho^A_{-\infty}) = & \frac{1}{2}\left[1-\nu_B\sin\left[2\Delta(f_A,f_B)\right]\left({}_A\langle 0|\rho^A_{-\infty}|0\rangle_A - {}_A\langle 1|\rho^A_{-\infty}|1\rangle_A\right)\right]|+\rangle_{B\,B}\langle +| \\
& + \frac{1}{2}\left[1+\nu_B\sin\sin\left[2\Delta(f_A,f_B)\right]\left({}_A\langle 0|\rho^A_{-\infty}|0\rangle_A - {}_A\langle 1|\rho^A_{-\infty}|1\rangle_A\right)\right]|-\rangle_{B\,B}\langle -| \\
& + \left(\frac{i\nu_B}{2}\cos[2\Delta(f_A,f_B)]\,|+\rangle_{B\,B}\langle -| + H.c\right).
\end{aligned} \tag{4.87}$$

Using

$$|0\rangle_A = \begin{pmatrix}1\\0\end{pmatrix}, \tag{4.88}$$

$$|1\rangle_A = \begin{pmatrix}0\\1\end{pmatrix}, \tag{4.89}$$

such that $\langle\sigma^z_A\rangle_{\rho^A_{-\infty}} \equiv \mathrm{Tr}\{\sigma^z_A\rho^A_{-\infty}\} = {}_A\langle 0|\rho^A_{-\infty}|0\rangle_A - {}_A\langle 1|\rho^A_{-\infty}|1\rangle_A$, the map which describes the quantum communication channel takes the form [11, 13]

$$\begin{aligned}\rho^B \equiv \mathcal{E}(\rho^A_{-\infty}) = \frac{1}{2}\left[1 - \nu_B \sin[2\Delta(f_A, f_B)]\, \langle\sigma^z_A\rangle_{\rho^A_{-\infty}}\right] |+\rangle_{B\,B}\langle+|+\\ + \frac{1}{2}\left[1 + \nu_B \sin[2\Delta(f_A, f_B)]\, \langle\sigma^z_A\rangle_{\rho^A_{-\infty}}\right] |-\rangle_{B\,B}\langle-|+\\ + \left(\frac{i\nu_B}{2}\cos[2\Delta(f_A, f_B)]\, |+\rangle_{B\,B}\langle-| + H.c\right).\end{aligned} \tag{4.90}$$

Now, with the communication channel settled, we can compute its capacity, i.e., the maximum rate at which the communication between Alice and Bob can be reliably done. In the next section we will discuss the case where Alice wants to transmit classical information to Bob by making use of the quantum communication channel described here and, also, calculate its classical channel capacity.

4.2 Classical Channel Capacity

4.2.1 Communication Procedure

Communication between Alice and Bob will take place in a very similar way to what has already been presented in previous chapters or, to be more precisely, in the second half of chapter 2. But now we will use some elements that were introduced in this chapter.

Therefore, suppose that Alice and Bob will communicate with each other by using a noisy channel $\mathcal{E}$, which is described according to Eq. (4.90). For the same reasoning as in Chapter 2, i.e., to allow the asymptotic behavior of the theory and minimize the error probability, we will consider that Alice makes several uses of the quantum channel.

The communication process will occur in the following way. Alice selects a message m from a set $X = 1, ..., |X|$, of size $|X|$. This message is encoded, through an encoder J, into a codeword $x^n(m)$, where n is the number of times she uses the communication channel $\mathcal{E}$. To each message, m, she associates a quantum state $\rho_m^{A_n}$ which is defined in the Hilbert space $\mathcal{H}_A^{\otimes n}$.

Since they are using a noisy channel, Bob receives a corrupted sequence y^n, which is associated to the random variable Y^n. In order to decode the message, he chooses a decoder $\mathcal{D}$, which is a positive-operator valued measure (POVM), i.e., $\mathcal{D} = \{F_{\hat{m}}^{B_n} | \hat{m} \in Y\}$. As Alice, he will associate a quantum state $\mathcal{E}(\rho_m^{A_n})$, defined in $\mathcal{H}_B^{\otimes n}$, to each received message.

A POVM is defined as a collection of positive operators in a Hilbert space that sum to the identity, for example, the operators $M_\mu^\dagger M_\mu$ used before, constitutes a POVM, since they are positive and satisfy

$$\sum_\mu M_\mu^\dagger M_\mu = I. \tag{4.91}$$

The POVM $F_{\hat{m}}^{B_n}$ is associated with the measurement of the outcome m, such that, the probability of obtaining it is given by [11]

$$P(Y = m|X = m) = \mathrm{Tr}\{F_m^{B_n}\mathcal{E}^{\otimes n}(\rho_m^{A_n})\}, \tag{4.92}$$

where, as we saw before, $P(Y = m|X = m)$ is the conditional probability of the random variable Y to carry the value m when $X = m$.

Consequently, the error probability must be

$$\begin{aligned} p_e(m, C) &= 1 - P(Y = m|X = m) && (4.93)\\ &= \mathrm{Tr}\{\mathcal{E}^{\otimes n}(\rho_m^{A_n}) - F_m^{B_n}\mathcal{E}^{\otimes n}(\rho_m^{A_n})\} && (4.94)\\ &= \mathrm{Tr}\{(I - F_m^{B_n})\mathcal{E}^{\otimes n}(\rho_m^{A_n})\}. && (4.95)\end{aligned}$$

where $C \equiv x^n(m)_{m\in X}$. Such that, the average error probability of this coding scheme is defined as

$$\bar{p}_e \equiv \frac{1}{|X|}\sum_{m=1}^{|X|} p_e(m, C), \tag{4.96}$$

Also, the maximum probability of error is

$$p_e^* \equiv \max p_e(m, C). \tag{4.97}$$

If $|X|$ is the size of the messages set, then the number of message bits must be $\log_2(|X|)$, also, with n uses of the communication channel, the rate of communication, for a coding scheme C, will be

$$R_C \equiv \frac{1}{n}\log_2(|X|). \tag{4.98}$$

According to Shannon theorems, the rate R_C is achievable if there's a code $(n, R_C - \Delta, \epsilon)$, with $\Delta > 0$ and the error $\epsilon > 0$, such that $p_e^* \leq \epsilon$, as it was already demonstrated in Chapter 2. Also, according to Shannon theorem in Chapter 2, the maximal achievable rate of communication is the channel capacity. In the following, we will calculate the channel capacity, which is denoted by $C(\mathcal{E})$, of the quantum channel $\mathcal{E}$.

4.2.2 The Channel Capacity

As discussed before, the amount of reliable information that Bob can get from Alice, for one use of the quantum channel, is limited by the channel capacity. Now we are going to calculate this limit by considering that Alice and Bob are communicating with each other through the channel $\mathcal{E}(\rho^A_{-\infty m})$, where, again, m is a message chosen by Alice, such that $m \in X$.

Therefore, considering the probability distribution p_m associated with the message m, we can use the *Holevo - Schumacher - Westmorland theorem*, stated as follows [3, 5, 11]:

Theorem 4. *The classical capacity of a quantum channel is equal to the regularization of the Holevo quantity of the channel:*

$$C(\mathcal{E}) = \chi_{reg}(\mathcal{E}) \tag{4.99}$$

where

$$\chi_{reg}(\mathcal{E}) \equiv \lim_{n\to\infty} \frac{1}{n}\chi(\mathcal{E}^{\otimes n}). \tag{4.100}$$

The $\chi(\mathcal{E})$ is called the *Holevo quantity* of the quantum channel $\mathcal{E}$, which is defined as an upper bound on the accessible information, i.e., the maximum amount of information that can be extracted from a quantum system. Therefore, if $I(X:Y)$ is the mutual information between Alice and Bob, where X is the random variable associated with the input and Y is the one associated with the output, the maximization of it must be [2, 3, 11]

$$I(X:Y) \leq \chi(\mathcal{E}) \equiv \max_{\{p_i,\rho_i\}} \left\{ S(\rho) - \sum_{i=1}^{|X|} p_i S(\rho_i) \right\}, \tag{4.101}$$

where $\rho = \sum_i p_i \rho_i$ and S is the *von Neumann entropy*. In communication theory, as we saw before, an entropy describes the average ignorance —or surprise— we have about the system. The von Neumann entropy is a generalization of the Shannon entropy, which we described above, for a quantum system. Moreover, if the Holevo information of the channel is additive, then $\chi_{reg}(\mathcal{E}) = \chi(\mathcal{E})$, this proof can be found in [3].

Thus, the channel capacity, $C(\mathcal{E})$, will be

$$C(\mathcal{E}) = \lim_{n\to\infty} \frac{1}{n}\chi(\mathcal{E}^{\otimes n}), \tag{4.102}$$

where n is the number of uses of the channel and $\chi(\mathcal{E}^{\otimes n})$ is the Holevo quantity, such that [11]

$$\chi(\mathcal{E}(\rho^A_{-\infty m})) \equiv \max_{\{p_m,\rho^A_{-\infty m}\}} \left\{ S\left(\rho^B_{\text{av}}\right) - \sum_m p_m S(\rho^B_m) \right\}, \tag{4.103}$$

where ρ^B_{av} is the average measured value, or the expectation valued, so that

$$\rho^B_{\text{av}} \equiv \mathcal{E}\left(\sum_m p_m \rho^A_m\right), \tag{4.104}$$

and $S(\rho)$ is the von Neumann entropy of the density matrix ρ, which is defined by

$$S(\rho) \equiv \text{Tr}\{-\rho \log_2 \rho\}. \tag{4.105}$$

Thus, the Holevo quantity will be

$$\chi(\mathcal{E}(\rho^A_{-\infty m})) = \max_{\{p_m,\rho^A_{-\infty m}\}} \left\{ S\left(\mathcal{E}\left[\sum_m p_m \rho^A_{-\infty m}\right]\right) - \sum_m p_m S(\mathcal{E}[\rho^A_{-\infty m}]) \right\}. \tag{4.106}$$

Assuming this information is additive, such that $\chi_{reg} = \chi$, we can write the classical channel capacity as [11]

$$C(\mathcal{E}(\rho^A_{-\infty m})) = \max_{\{p_m,\rho^A_{-\infty m}\}} \left\{ S\left(\mathcal{E}\left[\rho^A_{-\infty}\right]\right) - \sum_m p_m S(\mathcal{E}[\rho^A_{-\infty\, m}]) \right\}, \tag{4.107}$$

where

$$\rho^A_{-\infty} \equiv \sum_m p_m \rho^A_{-\infty\, m}. \tag{4.108}$$

The idea here is to find the eigenvalues of the maps $\mathcal{E}(\rho^A_{-\infty})$ and $\mathcal{E}(\rho^A_{-\infty\, m})$, such that, the maximization of the accessible information will be accomplished when $S(\mathcal{E}[\rho^A_{-\infty}])$ is maximum and $S(\mathcal{E}[\rho^A_{-\infty\, m}])$ is minimum.

In order to maximize the accessible information, we can write Alice's density matrices, $\rho^A_{-\infty\, m}$, by using the identity, $\mathbb{I}$, and its Bloch vectors, $\mathbf{r}_m \equiv (x_m, y_m, z_m)$, as [5, 11]

$$\rho^A_{-\infty\, m} = \frac{1}{2}\left(\mathbb{I} + \mathbf{r}_m \cdot \sigma_A\right), \tag{4.109}$$

where σ_A are the Pauli matrices associated to Alice's qubit. By doing this, we can use the above definition of $\rho^A_{-\infty}$ to calculate the expectation value $\langle\sigma^z_A\rangle_{\rho^A_{-\infty}}$, which appears in Eq. (4.90). Relative to $\rho^A_{-\infty\, m}$, it becomes

$$\begin{aligned}
\langle\sigma^z_A\rangle_{\rho^A_{-\infty\, m}} &\equiv \mathrm{Tr}\left\{\sigma^z_A\rho^A_{-\infty\, m}\right\} && (4.110)\\
&= \frac{1}{2}\,\mathrm{Tr}\{\sigma^z_A(\mathbb{I}+\mathbf{r}_m\cdot\sigma_A)\} && (4.111)\\
&= z_m. && (4.112)
\end{aligned}$$

Therefore, the quantum map in Eq. (4.90) will be

$$\begin{aligned}
\mathcal{E}(\rho^A_{-\infty\, m}) = {} & \frac{1}{2}\left[1 - z_m\nu_B\sin[2\Delta(f_A,f_B)]\right]|+\rangle_{B\,B}\langle+|+\\
&+\frac{1}{2}\left[1 + z_m\nu_B\sin[2\Delta(f_A,f_B)]\right]|-\rangle_{B\,B}\langle-|+\\
&+\left(\frac{i\nu_B}{2}\cos[2\Delta(f_A,f_B)]\,|+\rangle_{B\,B}\langle-|+H.c\right).
\end{aligned} \tag{4.113}$$

Also, we can repeat the same procedure in order to find the expectation value $\langle\sigma^z_A\rangle_{\rho^A_{-\infty}}$ and the quantum map $\mathcal{E}(\rho^A_{-\infty})$, relative to $\rho^A_{-\infty}$. By doing this, we get

$$\langle\sigma^z_A\rangle_{\rho^A_{-\infty}} = z \tag{4.114}$$

and

$$\begin{aligned}
\mathcal{E}(\rho^A_{-\infty}) = {} & \frac{1}{2}\left[1 - z\nu_B\sin[2\Delta(f_A,f_B)]\right]|+\rangle_{B\,B}\langle+|+\\
&+\frac{1}{2}\left[1 + z\nu_B\sin[2\Delta(f_A,f_B)]\right]|-\rangle_{B\,B}\langle-|+\\
&+\left(\frac{i\nu_B}{2}\cos[2\Delta(f_A,f_B)]\,|+\rangle_{B\,B}\langle-|+H.c\right),
\end{aligned} \tag{4.115}$$

where $z \equiv \sum_m p_m z_m$ [11].

Now, if we want to find the eigenvalues of these maps, we need to diagonalize it. To simplify the notation, let us make the following substitution:

$$\begin{aligned}
a &= z_m\nu_B\sin\left[2\Delta(f_A,f_B)\right]; && (4.116)\\
b &= \nu_B\cos\left[2\Delta(f_A,f_B)\right]. && (4.117)
\end{aligned}$$

The diagonalization of the matrix which represents the map $\mathcal{E}(\rho^A_{-\infty\, m})$ will be

$$det\begin{pmatrix} \frac{1}{2}-\frac{1}{2}z_m a-\lambda & \frac{i}{2}b \\ -\frac{i}{2}b & \frac{1}{2}+\frac{1}{2}z_m a-\lambda \end{pmatrix} = \left(\frac{1}{2}-\frac{1}{2}z_m a-\lambda\right)\times \\ \times\left(\frac{1}{2}+\frac{1}{2}z_m a-\lambda\right)-\frac{b^2}{2}=0, \tag{4.118}$$

thus, we find the quadratic equation

$$\lambda^2-\lambda-\frac{1}{4}(z_m^2a^2+b^2-1)=0. \tag{4.119}$$

By solving it we will get the eigenvalues of $\mathcal{E}(\rho^A_{-\infty\, m})$, which are [11]

$$\lambda_1=\frac{1}{2}+\frac{\nu_B}{2}\sqrt{z_m^2\sin^2[2\Delta(f_A,f_B)]+\cos^2[2\Delta(f_A,f_B)]}, \tag{4.120}$$

and

$$\lambda_2=\frac{1}{2}-\frac{\nu_B}{2}\sqrt{z_m^2\sin^2[2\Delta(f_A,f_B)]+\cos^2[2\Delta(f_A,f_B)]}=1-\lambda_1. \tag{4.121}$$

Again, the same procedure can be done for $\mathcal{E}(\rho^A_{-\infty})$, so we get the eigenvalues λ'_1 and $1-\lambda'_1$, where

$$\lambda'_1=\frac{1}{2}+\frac{\nu_B}{2}\sqrt{z^2\sin^2[2\Delta(f_A,f_B)]+\cos^2[2\Delta(f_A,f_B)]}. \tag{4.122}$$

Note that, since Alice and Bob are communicating with each other by using classical information, the accessible information can be written in terms of the classical Shannon entropy, $H(x)$, such as [11]

$$S(\mathcal{E}[\rho^A_{-\infty}])-\sum_m p_m S(\mathcal{E}[\rho^A_{-\infty\, m}])=\sum_i\left\{H(\lambda'_i)-\sum_m p_m H(\lambda_i)\right\}, \tag{4.123}$$

with $i=1,2$. Also,

$$H(\lambda_1)=-\lambda_1\log_2\lambda_1, \tag{4.124}$$

$$H(\lambda_2)=H(1-\lambda_1)=-(1-\lambda_1)\log_2(1-\lambda_1), \tag{4.125}$$

$$H(\lambda'_1)=-\lambda'_1\log_2\lambda'_1 \tag{4.126}$$

and

$$H(\lambda'_2)=H(1-\lambda'_1)=-(1-\lambda'_1)\log_2(1-\lambda'_1). \tag{4.127}$$

Or, in a more simplified way, we can just define

$$H(x) \equiv -x\log_2 x - (1-x)\log_2(1-x), \tag{4.128}$$

and write the accessible information as

$$I_{acc} = H(\lambda'_1) - \sum_m p_m H(\lambda_1), \tag{4.129}$$

where, according to Eqs. (4.120) and (4.122), $\lambda_1 \geq 1/2$ and $\lambda'_1 \geq 1/2$.

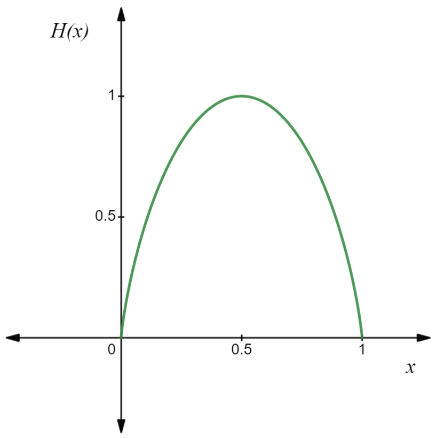

Figure 4.1: Dependence of the Shannon entropy $H(x)$, defined in Eq. (4.128), with x to $x \in [0, 1]$.

If we plot the graph of $H(x) \times x$, as in Fig. 4.1, it will show us that $H(x)$ is an open downward parabola with maximum point in $x = 1/2$, such that it always decrease for $x \geq 1/2$. Therefore, the maximum value of λ_1, and the minimum value of λ'_1, are, respectively,

$$\lambda_1 \leq \frac{1}{2} + \frac{\nu_B}{2} \tag{4.130}$$

and

$$\lambda'_1 \geq \frac{1}{2} + \frac{\nu_B}{2}|\cos[2\Delta(f_A, f_B)]|, \tag{4.131}$$

which were calculated by making $z_m^2 = 1$ and $z = 0$, that are the maximum and minimum values of z_m and z, respectively. Consequently, the entropies associated to these values can be written as [11]

$$H(\lambda_1) \geq H\left(\frac{1}{2} + \frac{\nu_B}{2}\right) \tag{4.132}$$

and

$$H(\lambda'_1) \leq H\left(\frac{1}{2}+\frac{\nu_B}{2}|\cos[2\Delta(f_A,f_B)]|\right). \tag{4.133}$$

Thus, with the maximum value of $H(\lambda'_1)$ and the minimum value of $H(\lambda_1)$, the maximization of the accessible information will be

$$I_{acc} \leq H\left(\frac{1}{2}+\frac{\nu_B}{2}|\cos[2\Delta(f_A,f_B)]|\right) - \sum_m p_m H\left(\frac{1}{2}+\frac{\nu_B}{2}\right), \tag{4.134}$$

such that, the Channel Capacity is [11, 13]

$$C(\mathcal{E}) = H\left(\frac{1}{2}+\frac{\nu_B}{2}|\cos[2\Delta(f_A,f_B)]|\right) - H\left(\frac{1}{2}+\frac{\nu_B}{2}\right). \tag{4.135}$$

Equation (4.135) has the interesting property that, if Alice and Bob are causally disconnected, then $\Delta(f_A, f_B) = 0$ and, consequently, $C(\mathcal{E}) = 0$. Which must be true, since Alice's messages can not reach Bob's detectors, as he is not in her causal future. But, if they are causal related, then $\Delta(f_A, f_B) \neq 0$ and $C(\mathcal{E}) \neq 0$, so that Bob and Alice can communicate.

Knowing only the channel capacity is not enough for Alice and Bob, they also need to know how much energy will be spent during the communication process. In the next section we will develop a quantum field theory restricted to null three-dimensional sub-manifolds, which we call cosmological horizon. This will be done with the help of the Penrose diagram, which will allow us to investigate how the total energy changes if we evaluate it from the asymptotic past to the asymptotic future.

4.3 QFT Restricted to Null Sub-Manifolds

In Chapter 3 we introduced the Globally Hyperbolic structure $(\mathbb{M}, g_{\mu\nu})$, where $\mathbb{M}$ is a four-dimensional manifold and $g_{\mu\nu}$ is the Lorentzian metric tensor, and we formulate a quantum theory of the Klein-Gordon scalar field ϕ based on the symplectic space $(\mathcal{S}^{\mathbb{C}}, \hat{\Omega})$, where $\mathcal{S}^{\mathbb{C}}$ is the space of complex solutions of the KG equation (3.36) and $\hat{\Omega}$ represents the fundamental observables.

But, since we are considering a curved spacetime, we need to deal with the possibility of having a black hole event horizon, or some other complex structure, on it. Therefore, let us consider the horizon $\mathfrak{h}$, which is a three-dimensional null hypersurface composed by a future null infinity (I^+) —lightlike surfaces that represents the endpoint of light's world-lines— and a event horizon $\mathfrak{h}^+$ of a black-hole, so that $\mathfrak{h} \equiv I^+ \cup \mathfrak{h}^+$. Also, assuming the asymptotic flatness of spacetime infinitely far away from the black hole, we can associate the QFT which will be presented here with the one we described in the last chapter.

There is a special class of diagram which will be very useful for us. Since we want to work with a spacetime with contains a causal horizon, we need to understand how its causal structure can be represented. The diagram which describes an infinite spacetime in a finite size is called a Penrose Diagram. Additionally, it preserves angles, such that light-cones will have the same slope as in Fig. 3.2 and, locally, the metric on it is conformally equivalent to the real one in spacetime.

The Penrose diagram can be represented for Minkowski spacetime as well as for Schwarzschild's or others, but since the construction uses the same technique we will illustrate it by using the Minkowski metric, which is described in Eq. (3.1). Thus, let us first write it in polar coordinates, as

$$ds^2 = -dt^2 + dr^2 + r^2\, d\Omega^2, \tag{4.136}$$

where $\Omega = d\theta^2 + \sin^2\theta d\phi^2$. Note that, this Ω is not the same as the bilinear map presented in Chapter 3.

Now, by defining the null coordinates u and v as

$$u = t - r \tag{4.137}$$

and

$$v = t + r, \tag{4.138}$$

with the limits $-\infty \leq u \leq \infty$ and $-\infty \leq v \leq \infty$, respectively, we can rewrite the Eq. (4.136) as [25]

$$ds^2 = -\frac{1}{2}(du \otimes dv + dv \otimes du) + \frac{1}{4}(v-u)^2 d\Omega^2. \tag{4.139}$$

In order to replace the infinite limits of the above equation we can use some function that goes to a fixed value when its argument goes to infinity. Fortunately, it is the case for $\arctan(x)$. Then, we can perform the change of coordinates according to [25]

$$\begin{cases} U = \arctan(u); \\ V = \arctan(v). \end{cases} \tag{4.140}$$

Where, $-\pi/2 < U < \pi/2$ and $-\pi/2 < V < \pi/2$. Therefore, using that $\sec(x) = 1/\cos(x)$, we have

$$du \otimes dv + dv \otimes du = \frac{1}{\cos^2 U \ \cos^2 V}(dU \otimes dV + dV \otimes dU) \tag{4.141}$$

and

$$\begin{aligned}
(v-u)^2 &= \tan^2(V) + \tan^2(U) - 2\tan(V)\tan(U) && (4.142)\\
&= \frac{\sin^2(V)}{\cos^2(V)} + \frac{\sin^2(U)}{\cos^2(U)} - 2\frac{\sin(V)}{\cos(V)}\frac{\sin(U)}{\cos(U)} && (4.143)\\
&= \frac{1}{\cos^2(U)\cos^2(V)}\left[\sin(V)\cos(U) - \sin(U)\cos(V)\right]^2 && (4.144)\\
&= \frac{\sin^2(V-U)}{\cos^2(U)\cos^2(V)}. && (4.145)
\end{aligned}$$

In these new coordinates, the metric takes the form [25]

$$ds^2 = \frac{1}{4\cos^2(U)\cos^2(V)}\left[-2(dU \otimes dV + dV \otimes dU) + \sin^2(V-U)d\Omega^2\right]. \tag{4.146}$$

But it can be simplified if we go back to timelike and radial coordinates. Thus, let us define

$$T = V + U \tag{4.147}$$

as the conformal timelike coordinate, and

$$R = V - U \tag{4.148}$$

as the conformal radial coordinate. In these equations, $-\pi < T < \pi$ and $0 < R < \pi$. Also, by making

$$\gamma^2 = 4\cos^2(U)\cos^2(V), \tag{4.149}$$

such that

$$\begin{aligned}
\gamma &= 2\cos(U)\cos(V) = [\cos(U-V) + \cos(U+V)] && (4.150)\\
&= \cos(R) + \cos(T) && (4.151)
\end{aligned}$$

we can write the new metric as [25]

$$ds^2 = \gamma^{-2}(T,R)(-dT^2 + dR^2 + \sin^2(R)d\Omega^2) = \gamma^{-2}(T,R)d\tilde{s}^2, \tag{4.152}$$

where $d\tilde{s}^2$ describes a manifold of topology $\mathbb{R} \times \mathbb{S}^3$, which can be defined as

$$d\tilde{s}^2 \equiv -dT^2 + dR^2 + \sin^2(R)d\Omega^2. \tag{4.153}$$

We can now present the Penrose diagram, which is a plot of $T \times R$, as illustrated in Fig. 4.2.The figure shows the future null infinite I^+, and the past null infinity I^-, which are null hypersurfaces with topology $\mathbb{R} \times \mathbb{S}^2$. $\mathbb{S}^2$ is a two-dimensional spacelike submanifold of Minkowski spacetime and $\mathbb{R}$ is the set of real numbers. This process is an illustration of the technique used to construct the Penrose diagram. Since we are working with a curved spacetime, this technique must be used to the specific metric associated to the manifold $\mathbb{M}$.

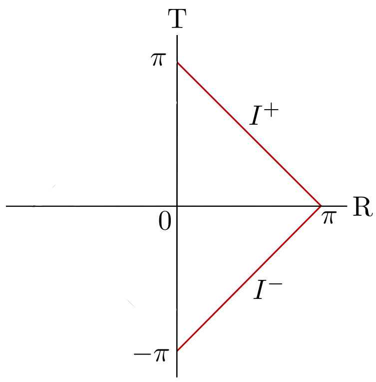

Figure 4.2: The Penrose diagram in Minkowski spacetime.

In a general way, we can consider the coordinates $(\sigma, \lambda, s^1, s^2)$ on $\mathbb{M}$, such that, when this process is performed, the restricted metric to $\mathfrak{h}$ will be [13]

$$g|_{\mathfrak{h}} = \gamma^2(-d\sigma \otimes d\lambda - d\lambda \otimes d\sigma + h_\Gamma), \tag{4.154}$$

where $\gamma^2 \in \mathbb{R}$. As before, the future and past null infinities have topology $\mathbb{R} \times \Gamma$, such that $\mathfrak{h}$ is diffeomorphic to it, and $\Gamma = \mathbb{S}^2$ is a two-dimensional spacelike submanifold of $\mathbb{M}$ with coordinates $s \equiv (s^1, s^2)$. Also, h_Γ is the induced metric in Γ and σ is a smooth function on $\mathbb{M}$, such that $\sigma|_{\mathfrak{h}} = 0$ and $d\sigma \neq 0$ everywhere on $\mathfrak{h}$ [13].

In this way, if ψ is a solution to the KG equation, we can define the complex space associated to it, on $\mathfrak{h}$, as [13]

$$\mathcal{S}^{\mathbb{C}}_{\mathfrak{h}} \equiv \{\psi : \mathfrak{h} \to \mathbb{C} | \psi, \partial_\lambda \psi \in L^2(\mathfrak{h}, d\lambda \wedge \epsilon_\Gamma)\}, \tag{4.155}$$

where $L^2(\mathfrak{h}, d\lambda \wedge \epsilon_\Gamma)$ is the space of square-integrable functions on $\mathfrak{h}$ and ϵ_Γ is the volume element in Γ. Also, we define the KG inner product between the functions ψ_1 and ψ_2, as

$$\langle \psi_1 | \psi_2 \rangle_{\mathfrak{h}} \equiv -i\Omega_{\mathfrak{h}}(\overline{\psi}_1, \psi_2), \tag{4.156}$$

where $\Omega_{\mathfrak{h}}$ is the symplectic product

$$\Omega_{\mathfrak{h}}(\psi_1, \psi_2) \equiv \int_{\mathfrak{h}} d\lambda \wedge \epsilon_\Gamma (\psi_2 \partial_\lambda \psi_1 - \psi_1 \partial_\lambda \psi_2), \tag{4.157}$$

with $\psi_1, \psi_2 \in \mathcal{S}^{\mathbb{C}}_{\mathfrak{h}}$ [13].

As we did in Chapter 3, we can use the projection operator K, to define the single-particle Hilbert space, i.e., the Hilbert space of all states that may be classified as one-particle states, as

$$\mathcal{H}_{\mathfrak{h}} \equiv \{K\psi | \psi \in \mathcal{S}^{\mathbb{C}}_{\mathfrak{h}}\}, \tag{4.158}$$

such that $K : \mathcal{S}^{\mathbb{C}}_{\mathfrak{h}} \to \mathcal{H}_{\mathfrak{h}}$. Again, it must satisfy the following properties:

i. $\mathcal{S}^{\mathbb{C}}_{\mathfrak{h}} = \mathcal{H}_{\mathfrak{h}} \oplus \overline{\mathcal{H}}_{\mathfrak{h}}$;

ii. the KG inner product is positive-definite in $\mathcal{H}_{\mathfrak{h}}$, such that $(\mathcal{H}_{\mathfrak{h}}, \langle | \rangle)$ is a Hilbert space;

iii. if $u \in \mathcal{H}_{\mathfrak{h}}$ and $v \in \overline{\mathcal{H}}_{\mathfrak{h}}$, then $\langle u | v \rangle_{\mathfrak{h}} = 0$.

If the Hilbert space comprehends all possible states of the quantum field, it can be defined as the symmetric Fock space $\mathcal{F}_S(\mathcal{H}_{\mathfrak{h}})$. On this space, the fundamental quantum observables, $\hat{\Omega}_{\mathfrak{h}}(\psi, \cdot)$, are described as

$$\hat{\Omega}_{\mathfrak{h}}(\psi, \cdot) \equiv ia(\overline{K\psi}) - ia^\dagger(K\psi), \tag{4.159}$$

where a and $a^\dagger$ are the creation and annihilation operators, respectively. Consequently, we can also define the field operator associated with a real solution ψ, as

$$\hat{\phi}^{(\mathfrak{h})}(\psi) \equiv \hat{\Omega}_{\mathfrak{h}}(\psi, \cdot) \equiv ia(\overline{K\psi}) - ia^\dagger(K\psi), \tag{4.160}$$

where, in this case, K is the map $K : \mathcal{S}_{\mathfrak{h}} \to \mathcal{H}_{\mathfrak{h}}$, and $\mathcal{S}_{(\mathfrak{h})}$ is the space of real solutions such that $\mathcal{S}_{\mathfrak{h}} \subset \mathcal{S}_{\mathfrak{h}}^{\mathbb{C}}$.

We can use Eq. (4.160) to write a relation between $\hat{\phi}^{(\mathfrak{h})}(\psi)$ and the quantum field operator $\hat{\phi}^{(\mathfrak{h})}(\lambda, s)$, that is defined as a function of the creation and annihilation operators, as well as $\{u_k\}$, which comprises a complete basis for the one-particle Hilbert space $\mathcal{H}_{\mathfrak{h}}$. Also, $\phi^{(\mathfrak{h})}(\lambda, s)$ satisfies the commutation relation [13]

$$[\hat{\phi}^{(\mathfrak{h})}(\lambda, s), \partial_\lambda \hat{\phi}^{(\mathfrak{h})}(\lambda', s')] = \frac{i}{2}\delta(\lambda - \lambda')\delta_\Gamma(s - s'). \tag{4.161}$$

Therefore, using Eq. (4.160), we can write

$$\begin{aligned} \hat{\phi}^{(\mathfrak{h})}(\psi) &= \hat{\Omega}_{\mathfrak{h}}[\psi, \hat{\phi}^{(\mathfrak{h})}(\lambda, s)] \\ &= \int_{\mathfrak{h}} d\lambda \wedge \epsilon_\Gamma[\hat{\phi}^{(\mathfrak{h})}\partial_\lambda\psi - \psi\partial_\lambda\hat{\phi}^{(\mathfrak{h})}] \\ &= 2\int_{\mathfrak{h}} d\lambda \wedge \epsilon_\Gamma(\partial_\lambda\psi)\hat{\phi}^{(\mathfrak{h})}(\lambda, s). \end{aligned} \tag{4.162}$$

This relation, together with (4.161), will be used in the next section, in order to calculate the energy cost for the information transmission.

There are many choices of Hilbert spaces which satisfies the above properties, such that it may generate many inequivalent Fock spaces [12]. We already faced that problem in Chapter 3. All we need to do is to formulate this theory by using the algebraic notation, since it will allow us to consider unitarily inequivalent states on an equal footing, i.e., without the need to make any specific choice.

Therefore, consider the asymptotically flat spacetime $(\mathbb{M}, g_{\mu\nu})$ with the horizon $\mathfrak{h} \equiv I^+ \cup \mathfrak{h}^+$, such that the algebraic space associated with $\mathfrak{h}$ will be $\mathcal{A}(\mathfrak{h}) \equiv \mathcal{A}(I^+) \otimes \mathcal{A}(\mathfrak{h}^+)$. If we assume a massless field, its world-lines will coincide with light's one, thus, its chronological past can be represented by the past null infinity I^-. Also, this region is infinitely far away from the horizon, as we can see by the Penrose diagram, which contains a singularity, in Fig. 4.3.

This assumption will be very useful for us because, since the region I^- is far away from the horizon, it will make spacetime to be asymptotically flat, such that the algebra of observables $\mathcal{A}(I^-)$

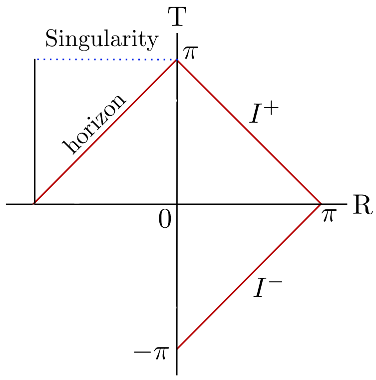

Figure 4.3: Penrose diagram with a singularity.

is the same presented on the last chapter, where the algebraic states are $\omega : \mathcal{A}(I^-) \rightarrow \mathbb{R}^+$ [11]. The interesting fact is that, as we are considering a massless field, the world-lines generated in I^- will end at $\mathfrak{h}$, or on I^+ if there is no singularity.

Because of this, we expect that all the information carried by be field will be transmitted to $\mathfrak{h}$ as well. In this way, we can associate the field operator, $\hat{\phi}(f)$, defined in the infinite null past, I^-, to the one we defined in Eq. (4.160), through the map [13]

$$\hat{\phi}(f) \rightarrow \hat{\phi}^{(\mathfrak{h})}(Ef^{\mathfrak{h}}), \tag{4.163}$$

where $f \in \mathcal{T}$. In this way, the algebraic states ω on $\mathcal{A}(I^-)$ give rise to the algebraic states $\omega_{\mathfrak{h}} : \mathcal{A}(\hat{\phi}(f)) \rightarrow \mathbb{R}^+$ on $\mathcal{A}(\mathfrak{h})$, thus, we can define [13]

$$\omega_{\mathfrak{h}}[\hat{\phi}^{(\mathfrak{h})}(Ef^{\mathfrak{h}})] \equiv \omega[\hat{\phi}(f)]. \tag{4.164}$$

Once we described the QFT for the globally hyperbolic spacetime $(\mathbb{M}, g_{\mu\nu})$, we are now prepared to calculate how much the total energy of the system changes when we evaluate it from the past null infinity, I^-, to the horizon $\mathfrak{h}$, which we defined to be $\mathfrak{h} \equiv \mathfrak{h}^+ \cup I^+$ if there is a singularity and $\mathfrak{h} \equiv I^+$ if there is not.

4.4 Energy Cost of the Communication Process

In the first half of the chapter we established the map which take us from Alice's qubit initial state, $\rho^A_{-\infty}$, to Bob's qubit final state, ρ^B, describing the communication channel. Then, it was used to compute the classical channel capacity, which is the maximum amount of reliable information Bob can get from Alice. But this is not enough, the transmission of information requires some energy cost and, to make sure if the communication is possible, they need to know how much energy this process will takes.

In this section we will consider the Globally Hyperbolic structure $(\mathbb{M}, g_{\mu\nu})$ such that we can foliate it with Cauchy surfaces Σ_t, from $\Sigma_{t\to-\infty} = I^-$ to $\Sigma_{t\to+\infty} = \mathfrak{h}$, where $I^-(\mathfrak{h})$ is the region far away the horizon $\mathfrak{h}$. Also, in order to investigate the energy cost of the transmission of information, we need to calculate the difference between the expectation values of the total energy of the system, on the initial and final times [13]. Therefore, given the system initial state

$$\rho_{-\infty} \equiv \rho^A_{-\infty} \otimes \rho^B_{-\infty} \otimes \rho_\omega, \tag{4.165}$$

the final state of the system will be

$$\rho_{+\infty} \equiv U\rho_{-\infty}U^\dagger = U(\rho^A_{-\infty} \otimes \rho^B_{-\infty} \otimes \rho_\omega)U^\dagger, \tag{4.166}$$

where U is the time evolution operator defined in Eq. (4.32).

If $H(t)$ is the Hamiltonian of the system, which is given in Eq. (4.1), the total energy variation of the system will be defined by [13]

$$\Delta E \equiv \langle H(+\infty)\rangle_{\rho_{+\infty}} - \langle H(-\infty)\rangle_{\rho_{-\infty}}. \tag{4.167}$$

But, as it was mentioned before, and according to Eq. (4.2), the interaction between the qubits with the field is sustained only for a finite amount of time Δt. In this way, we expect that, for $t \to \pm\infty$, the interaction Hamiltonian $H_{\text{int}}(t)$ goes to zero. Therefore, we just need to evaluate the energy considering the field Hamiltonian H_ϕ.

Moreover, let us note that $H_\phi(+\infty) = H^{\mathfrak{h}}_\phi$ and $\rho_{+\infty} = \rho^{\mathfrak{h}}_{+\infty}$ are the Hamiltonian of the field and the system state induced by $\rho_{-\infty}$ on $\mathfrak{h}$, respectively. Also, identifying $H(-\infty) = H^{(I^-)}_\phi$ and $\rho_{-\infty} = \rho^{(I^-)}_{-\infty}$, the total energy variation will be [13]

$$\begin{aligned} \Delta E &= \mathrm{Tr}\Big\{H^{(\mathfrak{h})}_\phi U^{(\mathfrak{h})} \rho^{\mathfrak{h}}_{-\infty} U^{(\mathfrak{h})\dagger}\Big\} - \mathrm{Tr}\Big\{H^{(I^-)}_\phi \rho^{I^-}_{-\infty}\Big\} && (4.168) \\ &= \mathrm{Tr}\Big\{U^{(\mathfrak{h})\dagger} H^{(\mathfrak{h})}_\phi U^{(\mathfrak{h})} \rho^{\mathfrak{h}}_{-\infty}\Big\} - \mathrm{Tr}\Big\{H^{(I^-)}_\phi \rho^{I^-}_{-\infty}\Big\}. && (4.169) \end{aligned}$$

where Eq. (4.166) was used to write the final system state.

In order to calculate energy contributions in Eq. (4.169) we need to define the field Hamiltonian on the null surfaces $\mathfrak{h}$ and I^-, as well as the evolution operator on $\mathfrak{h}$. The local energy and momentum of the field is described by the stress-energy tensor which, for the massless Klein-Gordon field, is

$$T^{\chi}_{\mu\nu} \equiv \boldsymbol{\nabla}_\mu \phi^\chi \boldsymbol{\nabla}_\nu \phi^\chi - \frac{1}{2} g_{\mu\nu} \boldsymbol{\nabla}_\gamma \phi^\chi \boldsymbol{\nabla}^\gamma \phi^\chi, \tag{4.170}$$

where $\chi = \mathfrak{h}, I^-$. We can use it to write the field Hamiltonian H^χ_ϕ as [13]

$$\begin{aligned} H^\chi_\phi &= \int_\chi d\lambda_\chi \wedge \epsilon_{\Gamma_\chi} T^\chi_{\mu\nu} k^\mu k^\nu & (4.171) \\ &= \int_\chi d\lambda_\chi \wedge \epsilon_{\Gamma_\chi} (\partial_{\lambda_\chi} \phi^\chi)^2, & (4.172) \end{aligned}$$

where $k^\mu \equiv (\partial_{\lambda_\chi})^\mu$ and $k^\nu \equiv (\partial_{\lambda_\chi})^\nu$ are the killing vectors tangent to the null generators of χ, i.e., the values of χ for constant (s^1, s^2) such that λ_χ describes null geodesics, and Γ_χ is a bi-dimensional spacelike surface transverse to the null generators of χ [13]. Lastly, we define the evolution operator as

$$U^{(\mathfrak{h})}_j \equiv e^{-i \sum_j \hat{\phi}^{(\mathfrak{h})}(E f^{\mathfrak{h}}_j) \otimes \sigma^z_j}, \tag{4.173}$$

where $j = A, B$. This is the only part of Eq. (4.32) which depends on the field operators and will not cancel.

Now, using the above definition of the evolution operator, we can evaluate the field Hamiltonian as

$$\begin{aligned} U^{(\mathfrak{h})\dagger} H^{(\mathfrak{h})}_\phi U^{(\mathfrak{h})} &= \int_{\mathfrak{h}} d\lambda_{\mathfrak{h}} \wedge \epsilon_{\Gamma_{\mathfrak{h}}} U^{(\mathfrak{h})\dagger} [\partial_{\lambda_{\mathfrak{h}}} \phi^{\mathfrak{h}}]^2 U^{(\mathfrak{h})} & (4.174) \\ &= \int_{\mathfrak{h}} d\lambda_{\mathfrak{h}} \wedge \epsilon_{\Gamma_{\mathfrak{h}}} (U^{(\mathfrak{h})\dagger} \partial_{\lambda_{\mathfrak{h}}} \phi^{\mathfrak{h}} U^{(\mathfrak{h})})^2. & (4.175) \end{aligned}$$

Therefore, using the relation

$$e^a \, b \, e^{-a} = b + [a, b], \tag{4.176}$$

which is valid when $[[a, b], a] = [[a, b], b] = 0$, and the evolution operator defined in Eq. (4.173), we can write

$$U_j^{(\natural)\dagger}\partial_{\lambda_\natural}\phi^{(\natural)}U_j^{(\natural)} = e^{i\sum_j \hat{\phi}^{(\natural)}(Ef_j^{\natural})\otimes\sigma_j^z}\partial_{\lambda_\natural}\phi^{(\natural)}e^{-i\sum_j \hat{\phi}^{(\natural)}(Ef_j^{\natural})\otimes\sigma_j^z} \tag{4.177}$$

$$= \partial_{\lambda_\natural}\phi^{(\natural)} + \left[i\sum_{j=A,B}\phi^{(\natural)}(Ef_j^{\natural})\otimes\sigma_j^z, \partial_{\lambda_\natural}\phi^{(\natural)}(\lambda'_\natural, s')\right]. \tag{4.178}$$

By equation (4.162), with $\psi = Ef_j^{\natural}$, we have

$$\hat{\phi}^{\natural}(Ef_j^{\natural}) = 2\int_{\natural} d\lambda_\natural \wedge \epsilon_{\Gamma_\natural}\partial_{\lambda_\natural}(Ef_j^{\natural})\hat{\phi}^{\natural}(\lambda_\natural, s), \tag{4.179}$$

such that

$$\begin{aligned}&[\phi^{\natural}(Ef_i^{\natural}), \partial_{\lambda_\natural}(Ef_j^{\natural})] = \\ &= \sum_{j=A,B}\left[2\int_{\natural} d\lambda_\natural \wedge \epsilon_{\Gamma_\natural}\partial_{\lambda_\natural}(Ef_j^{\natural})\phi^{\natural}(\lambda_\natural, s), \partial_{\lambda_\natural}\hat{\phi}^{(\natural)}(\lambda'_\natural, s')\right]\end{aligned} \tag{4.180}$$

$$= 2\sum_{j=A,B}\int_{\natural} d\lambda_\natural \wedge \epsilon_{\Gamma_\natural}\partial_{\lambda_\natural}(Ef_j^{\natural})\left[\phi^{(\natural)}(\lambda_\natural, s), \partial_{\lambda_\natural}\phi^{(\natural)}(\lambda'_\natural, s')\right] \tag{4.181}$$

$$= 2\sum_{j=A,B}\int_{\natural} d\lambda_\natural \wedge \epsilon_{\Gamma_\natural}\partial_{\lambda_\natural}(Ef_j^{\natural})\,\frac{i}{2}\delta(\lambda_\natural - \lambda'_\natural)\delta_{\Gamma_\natural}(s-s') \tag{4.182}$$

$$= \sum_{j=A,B} i\partial_{\lambda_\natural}Ef_j^{\natural}. \tag{4.183}$$

Now, we can substitute this result into Eq. (4.178), so we get

$$U_j^{(\natural)\dagger}\partial_{\lambda_\natural}\phi^{(\natural)}U_j^{(\natural)} = \partial_{\lambda_\natural}\phi^{\natural} - \sum_{j=A,B}\partial_{\lambda_\natural}(Ef_j^{\natural})\sigma_j^z. \tag{4.184}$$

In this way, the time evolution of $H_\phi^{(\natural)}$ will be,

$$U^{(\natural)\dagger}H_\phi^{(\natural)}U^{(\natural)} = \sum_{j=A,B}\int_{\natural} d\lambda_\natural \wedge \epsilon_{\Gamma_\natural}\left[\partial_{\lambda_\natural}\phi^{\natural} - \partial_{\lambda_\natural}(Ef_i^{\natural})\sigma_j^z\right]^2 \tag{4.185}$$

$$\begin{aligned}&= \int_{\natural} d\lambda_\natural \wedge \epsilon_{\Gamma_\natural}(\partial_{\lambda_\natural}\phi^{\natural})^2 - 2\sum_{j=A,B}\int_{\natural} d\lambda_\natural \wedge \epsilon_{\Gamma_\natural}(\partial_{\lambda_\natural}Ef_j^{\natural})(\partial_{\lambda_\natural}\phi^{\natural})\otimes\sigma_j^z + \\ &+ \sum_{i,j=A,B}\int_{\natural} d\lambda_\natural \wedge \epsilon_{\Gamma_\natural}(\partial_{\lambda_\natural}Ef_i^{\natural})(\partial_{\lambda_\natural}Ef_j^{\natural})\sigma_i^z\otimes\sigma_j^z,\end{aligned} \tag{4.186}$$

where the first integral is equal to $H_\phi^{(\natural)}$.

Thus, the total energy variation will be

$$\begin{aligned}\Delta E = {} & \operatorname{Tr}\Big\{H_\phi^{(\mathfrak{h})}\rho_{-\infty}^{\mathfrak{h}}\Big\} - \operatorname{Tr}\Big\{H_\phi^{I^-}\rho_{-\infty}^{(I^-)}\Big\} - \\ & - 2\operatorname{Tr}\left\{\rho_{-\infty}^{\mathfrak{h}}\sum_{j=A,B}\int_{\mathfrak{h}} d\lambda_{\mathfrak{h}}\wedge\epsilon_{\Gamma_{\mathfrak{h}}}(\partial_{\lambda_{\mathfrak{h}}}Ef_j^{\mathfrak{h}})(\partial_{\lambda_{\mathfrak{h}}}\hat{\phi}^{\mathfrak{h}})\otimes\sigma_j^z\right\} + \\ & + \operatorname{Tr}\left\{\rho_{-\infty}^{\mathfrak{h}}\sum_{i,j=A,B}\int_{\mathfrak{h}} d\lambda_{\mathfrak{h}}\wedge\epsilon_{\Gamma_{\mathfrak{h}}}(\partial_{\lambda_{\mathfrak{h}}}Ef_i^{\mathfrak{h}})(\partial_{\lambda_{\mathfrak{h}}}Ef_j^{\mathfrak{h}})\sigma_i^z\otimes\sigma_j^z\right\}\end{aligned} \tag{4.187}$$

$$\begin{aligned}\Delta E = {} & \operatorname{Tr}\Big\{H_\phi^{(\mathfrak{h})}\rho_{-\infty}^{\mathfrak{h}}\Big\} - \operatorname{Tr}\Big\{H_\phi^{I^-}\rho_{-\infty}^{(I^-)}\Big\} - \\ & - 2\sum_{j=A,B}\int_{\mathfrak{h}} d\lambda_{\mathfrak{h}}\wedge\epsilon_{\Gamma_{\mathfrak{h}}}(\partial_{\lambda_{\mathfrak{h}}}Ef_j^{\mathfrak{h}})\operatorname{Tr}\Big\{\rho_{-\infty}^{\mathfrak{h}}(\partial_{\lambda_{\mathfrak{h}}}\hat{\phi}^{\mathfrak{h}})\otimes\sigma_j^z\Big\} + \\ & + \sum_{i,j=A,B}\int_{\mathfrak{h}} d\lambda_{\mathfrak{h}}\wedge\epsilon_{\Gamma_{\mathfrak{h}}}(\partial_{\lambda_{\mathfrak{h}}}Ef_i^{\mathfrak{h}})(\partial_{\lambda_{\mathfrak{h}}}Ef_j^{\mathfrak{h}})\operatorname{Tr}\Big\{\rho_{-\infty}^{\mathfrak{h}}\sigma_i^z\otimes\sigma_j^z\Big\}.\end{aligned} \tag{4.188}$$

But we know, from Eq. (3.69), that $\hat{\phi}^{\mathfrak{h}}(\lambda_{\mathfrak{h}}, s)$ can be expanded as a function of the modes $\{u_k\} \subset \mathcal{H}^{\mathfrak{h}}$ and the operators $a(\overline{u_k})$ and $a^\dagger(u_k)$, i.e., the creation and annihilation operators, respectively. Thus, the expectation value $\left\langle \partial_{\lambda_{\mathfrak{h}}}\hat{\phi}^{\mathfrak{h}}\right\rangle_{\rho_{-\infty}^{\mathfrak{h}}} \equiv \operatorname{Tr}\Big\{\partial_{\lambda_{\mathfrak{h}}}\hat{\phi}^{\mathfrak{h}}\Big\}$ will be zero [13]. Also, if we separate the terms for which $j = i$ and $j \neq i$, respectively, the total energy variation will become

$$\begin{aligned}\Delta E = {} & \operatorname{Tr}\Big\{H_\phi^{(\mathfrak{h})}\rho_{-\infty}^{\mathfrak{h}}\Big\} - \operatorname{Tr}\Big\{H_\phi^{I^-}\rho_{-\infty}^{(I^-)}\Big\} + \int_{\mathfrak{h}} d\lambda_{\mathfrak{h}}\wedge\epsilon_{\Gamma_{\mathfrak{h}}}(\partial_{\lambda_{\mathfrak{h}}}Ef_i^{\mathfrak{h}})^2 + \\ & + 2\int_{\mathfrak{h}} d\lambda_{\mathfrak{h}}\wedge\epsilon_{\Gamma_{\mathfrak{h}}}(\partial_{\lambda_{\mathfrak{h}}}Ef_A^{\mathfrak{h}})(\partial_{\lambda_{\mathfrak{h}}}Ef_B^{\mathfrak{h}})\left\langle\sigma_A^z\right\rangle_{\rho_{-\infty}^A}\left\langle\sigma_B^z\right\rangle_{-\rho_{-\infty}^B}.\end{aligned} \tag{4.189}$$

We can separate the energy contributions into three parts. The first one will be [13]

$$E_\phi \equiv \operatorname{Tr}\Big\{H_\phi^{(\mathfrak{h})}\rho_{-\infty}^{\mathfrak{h}}\Big\} - \operatorname{Tr}\Big\{H_\phi^{I^-}\rho_{-\infty}^{(I^-)}\Big\}, \tag{4.190}$$

which depends only on the spacetime metric and the field state. This contribution is due to the particle creation when we go from I^- to $\mathfrak{h}$, coming from the change in the spacetime metric [13]. The second contribution is

$$E_j \equiv \int_{\mathfrak{h}} d\lambda_{\mathfrak{h}}\wedge\epsilon_{\Gamma_{\mathfrak{h}}}(\partial_{\lambda_{\mathfrak{h}}}Ef_j^{\mathfrak{h}})^2, \tag{4.191}$$

and, as we can see from Eq. (4.17), it depends on the coupling function of each qubit with the field, as well as their trajectories, so we can say that this contribution refers to the energy needed to switch on or off the qubits A and B, used by Alice and Bob, respectively [13]. Also, it is possible to show that the work necessary to switch on or off each qubit increases with the coupling strength, which describes the interaction of each qubit, separately, with the field.

Lastly, the third contribution will be

$$E_{AB} = 2\int_{\natural} d\lambda_{\natural} \wedge \epsilon_{\Gamma_{\natural}}(\partial_{\lambda_{\natural}} E f_A^{\natural})(\partial_{\lambda_{\natural}} E f_B^{\natural}) \langle \sigma_A^z \rangle_{\rho_{-\infty}^A} \langle \sigma_B^z \rangle_{\rho_{-\infty}^B}, \tag{4.192}$$

which describes an extra cost of the communication process [13]. Note that, it depends on each qubit initial state, such that, with the right choice of Bob's qubit initial state, since Alice's qubit initial state, $\rho_{-\infty}^A$, is fixed, we can make this contribution to vanish. The initial state which maximizes the signalling amplitude and, consequently, the channel capacity, was already chosen, it is described by Eq. (4.73). Therefore, after the initial states of each qubit are established, in such a way that the channel capacity is maximized, Alice and Bob can communicate with each other without any energy cost other than E_ϕ and E_j [13].

The Landauer principle says that, for any irreversible manipulation of information, such as erasing it, an amount of heat will be dissipated to the environment [14, 32]. As we will see in the next chapter, Landauer provides a fundamental limit of irreversible computation. By taking this principle into account, we can imagine a situation where Alice and Bob has already created their qubits for some purpose, expending an amount of energy $E_A + E_B$ to do that, the results of this chapter shows that no other energy cost will be needed in order to use these qubits to convey information. Therefore, we could think that Bob can uses the Landauer principle to erase the information —which was sent without any extra energy cost— and run a thermal machine, producing work. At first we could say that this process violates the second law of thermodynamics, since they are not expending any extra energy to convey information, but there is one thing we must considerate here. To erase the information, he also needs to spend some energy, which is equal to that dissipated on the environment after he erases the information, so the maximum work produced by the engine will be bounded by the energy he spent in the erasing process. Therefore, as we will see in more details on the next chapter, this procedure do not violates the second law of thermodynamics.

Chapter 5

Landauer Principle and the Quantum Communication Channel

In this chapter we are going to consider that Alice sent an amount of information to Bob equal to the channel capacity, which is described by Eq. (4.135), and we will use the Landauer principle to calculate the efficiency of a reversible heat engine, i.e., a mechanism that absorbs heat from some source and converts it partially into work in a cyclic process, in Bob's local frame. Additionally, to be consistent with the second law of thermodynamics, Bob must not produce more work than he spent to erase the information, so we will see that $W < |W_{er}|$, where W is the work produced by the heat engine and $|W_{er}|$ is the module of the work necessary to erase an amount of information equal to the channel capacity $C(\mathcal{E})$.

Since we already introduced the theory needed to understand the communication process, now we will need to investigate the Landauer principle and how it will be applied in Bob's local frame, which will be done on the first part of the chapter. After that we will present the Carnot cycle, so we can understand how the heat engine works, and, lastly, we will calculate the efficiency of this kind of engine to Bob's system.

5.1 Introduction to Landauer's Principle

In 1961 Landauer argued that an irreversible manipulation of information produces a certain amount of heat on the environment, which also increases its entropy by a finite quantity [14]. Therefore, if we manipulate one bit of information, so that the observer looses it, a minimal amount of heat, called the Landauer bound, must be produced.

An irreversible manipulation of information is a process that makes it impossible to discover the initial state of the system just based on its final state. As we said before, this kind of process can be done by erasing the information. To understand how we can erase the information about a system, let us consider the following example.

Suppose that our system is composed by a box with a movable partition and a molecule, as shown in Fig. 5.1. Here we want to erase the information about the position of the molecule. The molecule could be, initially, in any side of the box. Despite it is shown to be on the right side, we do not know that.

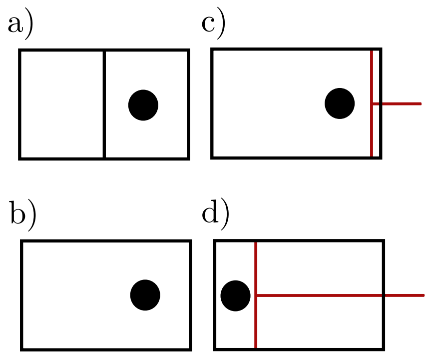

Figure 5.1: The process of erasing the information about a molecule's initial position. Fig. b) is the box without the partition, c) represents the box with a piston and d) is the process of moving it, changing the molecule's position.

The process of erasing the information about the molecule's position consists of three simple steps. First we remove the partition of the box. Then, we include a piston to it. Lastly, we use the piston to shift the molecule to the left side of the box. By doing that the molecule will be on the left side independently of its initial position. We say that this process erased the information because we have no way to find out what was the initial position of the molecule [16].

A connection between information and thermodynamics can be seen if we consider a case similar to that indicated by the last example, but now imagine that both sides of the box are full with molecules of a gas in some initial temperature T_0. An observer —commonly called Maxwell's demon— which has access to this experiment and is a sufficiently intelligent creature to know the positions and velocities of the molecules, can open a trap door on the partition and separate the molecules according with its velocities. He allows the faster ones to be in one side of the box and the slower ones in the other side. Molecules moving faster means that we will have more collisions with the walls of the box, increasing the pressure and temperature on this side.

According to thermodynamics, if we have a system with two different temperatures, we could run a heat engine and produce work. A heat engine produces work by performing a Carnot cycle, which will be explained on the next section. But the fact is, if the demon erases the information by removing the partition of the box, such that the system returns to its initial state, we could say that it is a cyclic process. In this way, the demon could keep creating this temperature gradient and run the heat engine just by knowing the position and velocities of the molecules, which, apparently, violates the second law of thermodynamics, since he is not paying any energetic cost to produce work. This paradox can be solved by noticing that the process of getting the information about the molecules is part of the full thermodynamic cycle. Therefore, since the demon is also part of the system, the information stored in his memory needs to be erased so that this process can be really cyclic [14, 16, 32, 33].

But how can we solve this problem by including the demon's memory to the erasure procedure? The first step we must take to understand how this problem can be solved is to consider the demon's mind as a physical system. If the demon removes the box partition, allowing the gas to return to its initial state, the whole system —which includes the demon and the box— will complete an apparent cycle, but it happens just because we assume that the information about the position and velocities of the molecules can be obtained reversibly without any energy cost [16]. However, when we consider the demon's mind as a physical system, which stores physical information, we need to erase it as well and, according to Landauer [14], this process does have a cost.

The Landauer's principle says that the energy cost to erase the information on the demon's mind is [14, 33]

$$W_{\text{er}} = -k_B T \ln 2, \tag{5.1}$$

per bit of information. Where k_B is the Boltzmann constant and T is the temperature of the system. Additionally, erasing the information is an irreversible process, therefore, still according to Landauer's principle —its experimental verification can be seen in [32]— this manipulation produces an amount of heat equivalent to

$$Q = k_B T \ln 2, \tag{5.2}$$

on the environment, per bit lost. If we consider that the engine operates at constant temperature, the extracted work will be

$$W_{\text{ex}} = Q = k_B T \ln 2. \tag{5.3}$$

In this way, we can calculate the total work gained by

$$W_{\text{total}} = W_{\text{er}} + W_{\text{ex}} = -k_B T \ln 2 + k_B T \ln 2 = 0. \tag{5.4}$$

Therefore, since all the energy produced by the engine must be used to erase the demon's memory, there are no violation of the second law of thermodynamics. What happened was that we thought that the work produced by the heat engine was the net work but this work is, actually, needed to erase the information on the demon's mind, so that no net work is produced [16].

By considering the demon's mind as a physical system, the information it stores is also physical and can play a role when defining the state of the system [16]. Moreover, the Landauer's principle is not only a solution of the Maxwell's demons paradox, but it establish a fundamental limit to irreversible manipulation of information. At the end of this chapter we will work with a similar experiment, where we will consider the case discussed before, when Alice sent an amount of information to Bob and he erases it in order to run a heat engine and produce work.

5.2 Heat Engine and Carnot Cycle

On the last section we introduced a basic notion of the Landauer's principle and how it can be used to solve the problem of Maxwell's demon. At this point the details about the operation the heat engine were not so important since we were focusing on learning the Landauer's fundamental limit, however, at the end of this chapter we want to use the principle to operate a heat engine on Bob's frame and calculate its efficiency. Therefore we need to understand how the engine works as well as its limitations.

As we saw from the last section, a heat engine is a mechanism that can perform work from heat cyclically or not. Here we will restrict ourselves to reversible processes, which means that the system as well as its vicinity can go back to its initial states always when the process is done. A reversible heat engine is also known as a Carnot engine.

The Kelvin' statement for the second law of thermodynamics is [34]:

Theorem 5. *It is impossible for any device operating cyclically to receive heat from a thermal reservoir and produce an equivalent amount of work.*

Therefore, there is no mechanism that can absorbs an amount of heat from the environment, produce an equivalent quantity of work and, at the end of the process, return to its initial state. There are always some losses on the way. To understand how it happens we will analyse the cycle of a Carnot engine, which is also known as the Carnot cycle.

An implication of the above discussion is that, at least two thermal reservoir, at different temperatures, are necessary in order to produce work by a heat engine operating cyclically. The reservoir with the highest temperature is the hot source, while the one with lower temperature is called cold sink. The engine absorbs an amount of heat from the hot source and produces work, but, in order to return to its initial state, the system will dissipate another amount of heat to the cold sink. This process will be better understood when we present the Carnot cycle.

According to Clausius' statement for the second law of thermodynamics [34–36]:

Theorem 6. *It is impossible to construct a device that operates in a cycle and produces no effect other than the transfer of heat from a cooler body to a hotter body.*

While Kelvin's statement implies that it's impossible to create a perfect engine, Clausius' statement suggests that there is no way to create a perfect refrigerator, i.e., a mechanism which extracts heat from the cold sink and provides it to the hot source, where both apparatus operates cyclically. But we know from thermodynamics that both statements are equivalents, since there are no miraculous heat engines or refrigerators.

Taking into account that an engine operating cyclically between two sources must obey these two statements, we need to construct a system where the heat conduction will only be made by bodies that are on the same temperature, so that it can be reversible. On the contrary, if we considered the flow of heat between bodies at different temperatures, the system could not return to the initial state without violating Clausius theorem. Consequently, the temperature variations of the system must occur without heat exchanges.

A cycle which satisfy these conditions is reversible and, as we said before, it is called Carnot cycle. Let us consider a system where we have an ideal gas stuck between adiabatic walls, i.e., walls that prevent the heat flow, a removable base and a piston that is localized over the gas. The Fig 5.2 illustrates the steps of a reversible cycle of this system, let us analyse it.

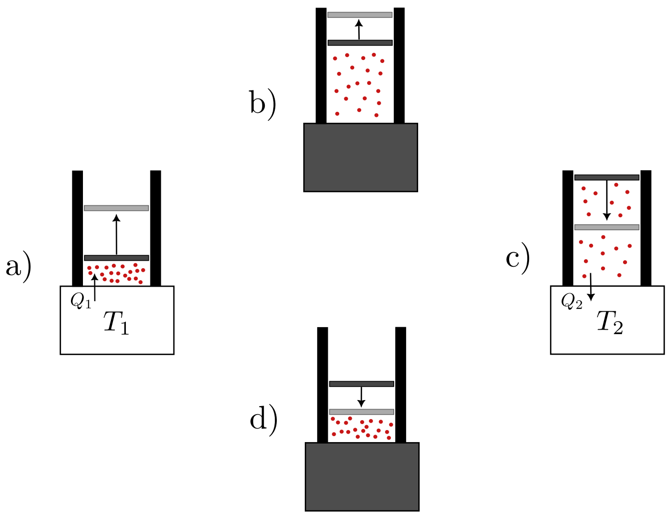

Figure 5.2: The Carnot Cycle. Where, a) describes a reversible isothermal expansion of the gas, b) is the reversible adiabatic expansion, c) represents the reversible isothermal compression and d) illustrates the reversible adiabatic compression. Note that there is only one piston on the system, which is the dark grey bar, the lighter one represents the final position of the piston.

The first step is represented by Fig 5.2.a). It is a reversible isothermal expansion of the gas. In this process, the gas absorbs an amount of heat equal to Q_1 from the hot source at temperature T_1, therefore, it expands isothermally and reversibly and does work. Although, we must remember that a Carnot cycle is an ideal process, i.e., it must be performed without any dissipation and in a quasi-static way so it can be reversible. Additionally, in a Carnot cycle, the isothermal heat transfer — the steps a) and c) of Fig. 5.2 — occurs by considering an infinitesimal gradient of temperature between the gas and the reservoir, the heat flows very slowly from the hotter system to the cooler system. This slow transfer allows the system to be brought back to its initial state by reversing the direction of the heat transfer, as it is done in a) and c), without any loss of energy.

The next step, represented by Fig 5.2.b), is the reversible adiabatic gas expansion. We place the system over an insulating base, it prevents the heat exchanges but the gas keeps expanding and doing work while its internal energy decreases, consequently, its temperature drops from T_1 to T_2 which enables us to go to the third step.

Now, the system will be prepared to return to its initial state, it is placed over a thermal reservoir at temperature $T_2 < T_1$, i.e., the cold sink. This process is called a reversible isothermal gas compression. The environment does work on the gas, which provides an amount of heat equal to Q_2 to the cold sink, as is illustrated by Fig. 5.2.c). On the last step we substitute the base by an insulating one and submits the gas to a reversible adiabatic compression. Compressing the gas will increase its temperature until it can reach T_1 and we can do all the process again. This completes the Carnot cycle.

For the next section we are interested in calculating the efficiency of a Carnot engine on Bob's inertial frame, e.g., its laboratory. A Carnot engine, operating between two thermal reservoirs at different temperatures, can be illustrated by Fig. 5.3. As already mentioned before, it absorbs an amount of heat Q_1 of the hot source, which is at temperature T_1, performing the work W and dissipating Q_2 on the cold sink.

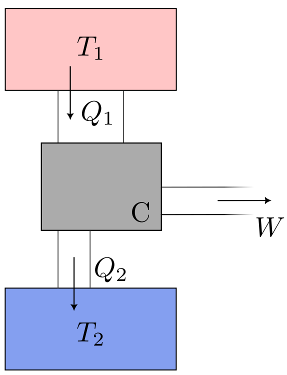

Figure 5.3: Representation of a Carnot engine C operating with two reservoirs.

The Carnot theorem says that the maximum effectiveness of a Carnot engine, operating between two thermal reservoirs, can be calculated by

$$\eta = \frac{W}{Q_1} = 1 - \frac{Q_2}{Q_1}, \tag{5.5}$$

where W is the total work performed by the system, i.e., $W = Q_1 - Q_2$, Q_1 is the heat absorbed from the hot source and Q_2 is the heat dissipated to the cold sink, as is shown by Fig. 5.3. In the next section, these concepts will be of great importance in order to calculate the work and the effectiveness of a Carnot engine on Bob's inertial frame.

5.3 Using Landauer Principle to Run a Carnot Engine

According to Landauer principle, to each bit of information erased an amount of heat equal to $k_B T \ln 2$ —which is about 3×10^{-21}J at 300K— is dissipated on the environment [14, 32]. Here we will consider the situation where Alice and Bob are using the quantum channel $\mathcal{E}$, which is described by Eq. (4.90), to convey information and, after this procedure, Bob will erase the received information, generating an amount of heat, in order to run a heat engine.

Therefore, according to the discussion of the last chapter, Bob must have at least two thermal reservoirs in order to run a heat engine cyclically. But, since we are considering a local frame, i.e., Bob's laboratory, these reservoirs cannot be too big. Hence, we can suppose that he has two identical reservoirs —which can be two water tanks— of mass m, at temperature T_0 and specific heat capacity c. He cannot run the heat engine if the temperature of the reservoirs are the same. In this way, the Landauer's principle will be of great importance so that Bob can run the engine.

Now we must apply the Landauer's principle to the system described in the last chapter and see what are the results of it. Consider that, after Alice and Bob's qubits are prepared, she sends the maximum amount of reliable information to Bob, with one use of the channel $\mathcal{E}$. Since we are using classical information, Bob receives an amount of bits equal to $C(\mathcal{E})$. Therefore, if he erases this information without reading it —he could read, but his memory would need to be erased too, so let us consider that he didn't read— an amount of heat will be dissipated to the environment. In this way, by using Landauer's principle, it will be

$$\begin{aligned} Q &= C(\mathcal{E})k_B T \ln 2 && (5.6) \\ &= \left[H\left(\frac{1}{2}+\frac{\nu_B}{2}|\cos\left[2\Delta(f_A, f_B)\right]|\right) - H\left(\frac{1}{2}+\frac{\nu_B}{2}\right)\right] k_B T \ln 2. && (5.7) \end{aligned}$$

Note that, since $C(\mathcal{E})$ is the maximum amount of information that Bob can erase per use of the channel, then Q will be the maximum amount of heat which will be produced by the erasing procedure. Additionally, if Alice and Bob are not causal related, then, no information can be erased and $Q = 0$.

From now on we will be focusing on Bob's inertial frame. The next step is to use the amount of heat Q, generated by erasing the message, to increase the temperature of one of the thermal reservoirs. If T_0 is the initial temperature of the reservoir 1, m is its mass and c its specific heat capacity, then, its final temperature T_1 can be calculated by

$$Q = mc(T_1 - T_0), \tag{5.8}$$

such that

$$T_1 = T_0 + \frac{Q}{mc}, \tag{5.9}$$

where Q is given by Eq. (5.7). Again, note that, if Alice and Bob are not causal related, the final temperature of the reservoir 1 will remain T_0, which implies that Bob couldn't run the heat engine. So, if there's no information transmitted through the channel, there will be no production of heat and, consequently, no temperature difference between the reservoirs, so the engine won't run.

Now we have two reservoirs at different temperatures, i.e., the reservoir 1 at T_1 and the reservoir 2 at T_0, so Bob can perform the Carnot cycle and calculate its efficiency. Considering the system described on the last section, where the heat engine is composed by an ideal gas situated between adiabatic walls with a removable base and a piston over it, as illustrated by Fig. 5.2, but instead of T_1 and T_2 we have T_1 and T_0 for the hot source and the cold sink, respectively.

Following the same procedure of the last section, but now on Bob's laboratory, the gas absorbs, reversibly, an amount of heat equal to Q_1 from the hot source and goes from the initial volume V_0

to the final one V_1. Thus, to the isothermal expansion of an ideal gas, this amount of heat will be

$$Q_1 = nRT_1 \ln\left(\frac{V_1}{V_0}\right), \tag{5.10}$$

where n is the number of moles of the gas and R is the universal gas constant $-R = 8.314\ J \cdot K^{-1} \cdot \text{mol}^{-1}-$ which, if we use Eq. (5.9), becomes

$$Q_1 = nR\left(T_0 + \frac{Q}{mc}\right)\ln\left(\frac{V_1}{V_0}\right), \tag{5.11}$$

with Q described by Eq. (5.7), m and c are the mass and the specific heat capacity of the reservoir, respectively. Additionally, the work produced by the gas when it goes from the state 0 to the state 1 will be [34]

$$W_{0\to1} = nR\left(T_0 + \frac{Q}{mc}\right)\ln\left(\frac{V_1}{V_0}\right). \tag{5.12}$$

Now, Bob substitutes the hot source for an insulating base and let the gas expands adiabatically from V_1 to V_2, so its temperature and pressure decreases from T_1 to T_0 and P_1 to P_2, respectively, such that [34]

$$V_1^{\gamma-1}T_1 = V_2^{\gamma-1}T_0 = V^{\gamma-1}T = C, \tag{5.13}$$

or we can write, by using the ideal gas state equation $P \propto T/V$,

$$P_1V_1^{\gamma} = P_2V_2^{\gamma} = PV^{\gamma} = C, \tag{5.14}$$

where γ is the ratio between the thermal capacities at constant pressure, C_P, and volume, C_V, of the gas, respectively, and C is a constant. If we consider a mono-atomic ideal gas γ will be equal to $5/3$. The work produced by the gas in the reversible adiabatic expansion will be

$$W_{1\to2} = -dU = \int_{V_1}^{V_2} PdV, \tag{5.15}$$

where dU is the variation of the internal energy of the gas and P is its pressure. By Eq. (5.14), we have

$$\begin{aligned} W_{1\to2} &= C\int_{V_1}^{V_2}\frac{dV}{V^{-\gamma}} &(5.16)\\ &= -\frac{(P_2V_2 - P_1V_1)}{\gamma-1}, &(5.17)\end{aligned}$$

but if Bob uses a mono-atomic ideal gas, we have

$$W_{1\to 2} = -\frac{3}{2}(P_2V_2 - P_1V_1). \tag{5.18}$$

Now the system needs to return to its initial state. Therefore, the gas will be put in contact with the cold sink, at temperature T_0, providing an amount of heat Q_2 to it and performing a reversible isothermal compression. The volume of the gas will decrease from V_2 to V_3, so the quantity Q_2 will be

$$Q_2 = nRT_0 \ln\left(\frac{V_3}{V_2}\right). \tag{5.19}$$

The internal energy variation ΔU of the gas will be zero, since its temperature remains constant during this process. Therefore, if we use the thermodynamic first law, we have that $Q_2 = W_{2\to 3}$, i.e., the amount of heat transferred to the cold sink is equal to the work done over the system. Note that, the convention we are using about W is that, $W > 0$ to the work performed by the system, additionally, $Q > 0$ represents always the heat provided to a system. In this way, the work done over the system will be

$$W_{2\to 3} = nRT_0 \ln\left(\frac{V_3}{V_2}\right). \tag{5.20}$$

The last step is to substitute the cold sink to an isolating base, so that the system returns to its initial state. In this process, the gas will undergo a reversible adiabatic compression, causing its temperature to increase until it reaches T_1. Here, there are no heat transfer between the system and the environment, thus, the internal energy of the gas will increase as Bob does work on the system, so it can returns to its initial state,

$$W_{3\to 0} = \Delta U = -\frac{3}{2}nR(T_1 - T_0). \tag{5.21}$$

The four stages of the Carnot cycle described above can be illustrated by the (P, V) diagram in Fig. 5.4.

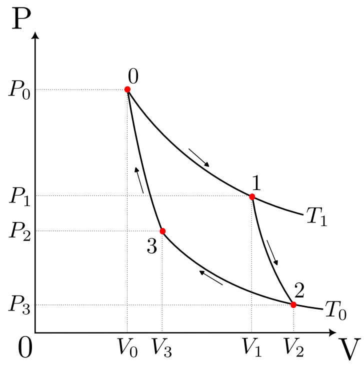

Figure 5.4: (P, V) diagram of the Carnot cycle.

Now, in order to calculate the maximum effectiveness of this engine, we can use the Eq. (5.5),

$$\eta = 1 - \frac{Q_2}{Q_1} = 1 - \frac{T_0}{T_1}, \tag{5.22}$$

where T_0 and T_1 are the temperatures of the cold and hot sources, respectively. If we use Eqs. (5.9) and (5.7), we get

$$\eta = 1 - \frac{T_0}{T_0 + \frac{Q}{mc}} \tag{5.23}$$

$$= 1 - T_0 \left[T_0 + \frac{C(\mathcal{E}) k_B T \ln 2}{mc} \right]^{-1}. \tag{5.24}$$

Note that this result depends on the causal relation between Alice and Bob. If they are not causally related, Bob will not be able to run the heat engine and its efficiency should be zero. In fact, this is what we get from Eq. (5.24), since $C(\mathcal{E})$ will be zero in this situation. But, if there is a causal relation between Alice and Bob, he can proceed with the steps explained in this Chapter and run the heat engine on his laboratory, such that $\eta \neq 0$. According to Carnot theorem Eq. (5.24) represents the maximum efficiency of a thermal engine operating cyclically, thus, we can calculate the maximum work by

$$W = \eta Q_1, \tag{5.25}$$

where Q_1 is the heat absorbed from the hot source. If the engine absorbs an amount of heat equal to Q, we will get

$$W = \eta Q = \left(1 - \frac{T_0}{T_1}\right) Q \tag{5.26}$$

$$= \left\{1 - T_0\left[T_0 + \frac{C(\mathcal{E})k_B T \ln 2}{mc}\right]^{-1}\right\} C(\mathcal{E})k_B T \ln 2. \tag{5.27}$$

Therefore, the work produced by the heat engine will always be smaller than the energy generated when the information was erased, satisfying the second law of thermodynamics.

5.4 Analysing the Channel Capacity to the Case of Inertial Detectors

In order to illustrate the results of the last chapters, let us consider an example in Minkowski spacetime. The spacetime is described by the inertial Cartesian coordinates $(t, x, y, z) \in \mathbb{R}^4$ and the Minkowski metric in Eq. (3.1), also, its causal structure is given by the light-cone, which can be seen in Fig. 3.2.

The Klein-Gordon equation for a massless field ϕ propagating on this spacetime will be

$$\Box\phi = 0, \tag{5.28}$$

where $\Box = -\sum_{\mu\nu} \eta^{\mu\nu}\partial_\mu\partial_\nu$, with $\eta^{\mu\nu}$ being the inverse of the components of the Minkowski metric tensor.

Therefore, by using this structure we can suppose the following situation: Alice is at rest relative to Bob and they are separated by a distance L. We can consider that she is at the origin of our inertial frame and Bob is at the position $\mathbf{x} = (L, 0, 0)$, where $\mathbf{x} \equiv (x, y, z)$ [13]. As it is discussed in chapter 4, Alice and Bob interacts their qubits with the field ϕ —which is used as the communication channel— in a limited amount of time, in order to convey information. We will use T_j^i and T_j^f, with $j = A, B$, as the initial and final interaction times, respectively. In this way, the function which describes Alice's qubit interaction with the field will be, according to Eq. (4.17),

$$f_A(t, \mathbf{x}) = \epsilon_A c_A(t)\delta^3(\mathbf{x}), \tag{5.29}$$

with ϵ_A being the dimensionless coupling constant and

$$c_A(t) = \begin{cases} e^{\alpha_A(t-T_A^i)}, & t < T_A^i \\ 1, & T_A^i \leq t \leq T_A^f \\ e^{-\alpha_A(t-T_A^f)}, & t > T_A^f \end{cases} \tag{5.30}$$

describes the switching function to the inertial time t —most of these calculations can be found in [13]. Note that, if $t \to -\infty$ then $c_j(t) \to 0$ and the same happens when $t \to +\infty$, so the information is lost by decoherence. For Bob's qubit interaction we can write

$$f_B(t, \mathbf{x}) = \epsilon_B c_B(t) \delta^3(\mathbf{x} - L\hat{\mathbf{x}}), \tag{5.31}$$

where ϵ_B is Bob's qubit dimensionless coupling constant and, again,

$$c_B(t) = \begin{cases} e^{\alpha_B(t-T_B^i)}, & t < T_B^i \\ 1, & T_B^i \leq t \leq T_B^f \\ e^{-\alpha_B(t-T_B^f)}, & t > T_B^f. \end{cases} \tag{5.32}$$

In order to analyse the channel capacity, we need to know $\Delta(f_A, f_B)$ and ν_B, since it depends on these quantities. Thus, let $\mathcal{S}$ be the space of real solutions, $Ef(\mathbf{x})$, where E is the map $E : \mathcal{T} \to \mathcal{S}$ —whose properties are shown in chapter 3. It is possible to show that the solutions Ef can be written as [37]

$$Ef(\mathbf{x}) = \int \epsilon_{\mathbb{M}} f(\mathbf{x'}) E(\mathbf{x}, \mathbf{x'}), \tag{5.33}$$

where

$$E(\mathbf{x}, \mathbf{x'}) \equiv \frac{1}{4\pi|\mathbf{x} - \mathbf{x'}|} \left[\delta(t - t' - |\mathbf{x} - \mathbf{x'}|) - \delta(t - t' + |\mathbf{x} - \mathbf{x'}|)\right]. \tag{5.34}$$

Now, if we use Eq. (5.31), we can write Ef_B as

$$\begin{aligned} Ef_B(t, \mathbf{x}) &= \int \epsilon_{\mathbb{M}} \frac{\epsilon_B C_B(t) \delta^3(\mathbf{x} - L\hat{\mathbf{x}})}{4\pi|\mathbf{x} - \mathbf{x'}|} [\delta(t - t' - |\mathbf{x} - \mathbf{x'}|) - \qquad (5.35) \\ &- \delta(t - t' + |\mathbf{x} - \mathbf{x'}|)] \\ &= \frac{\epsilon_B}{4\pi|\mathbf{x} - L\hat{\mathbf{x}}|} \left[c_B(t - |\mathbf{x} - L\hat{\mathbf{x}}|) - c_B(t + |\mathbf{x} - L\hat{\mathbf{x}}|)\right]. \qquad (5.36) \end{aligned}$$

We can substitute Eqs. (5.36) and (5.29) in Eq. (3.71), so that $\Delta(f_A, f_B)$ will be

$$\begin{aligned}\Delta(f_A, f_B) &\equiv \int \epsilon_{\mathbb{M}} f_A E f_B \qquad (5.37)\\ &= \int \epsilon_{\mathbb{M}}\, \epsilon_A C_A(t)\delta^3(\mathbf{x}) \frac{\epsilon_B}{4\pi|\mathbf{x}-L\hat{\mathbf{x}}|}\left[c_B(t-|\mathbf{x}-L\hat{\mathbf{x}}|)-\right.\\ &- \left. c_B(t+|\mathbf{x}-L\hat{\mathbf{x}}|)\right] \qquad (5.38)\\ &= \frac{\epsilon_A\epsilon_B}{4\pi}\int \epsilon_{\mathbb{M}} \frac{c_A(t)\delta^3(\mathbf{x})}{|\mathbf{x}-L\hat{\mathbf{x}}|}\left[c_B(t-|\mathbf{x}-L\hat{\mathbf{x}}|)-\right.\\ &- \left. c_B(t+|\mathbf{x}-L\hat{\mathbf{x}}|)\right] \qquad (5.39)\\ &= \frac{\epsilon_A\epsilon_B}{4\pi L}\int_{\mathbb{R}} dt\; c_A(t)[c_B(t-L)-c_B(t+L)]. \qquad (5.40)\end{aligned}$$

Moreover, we can define the one-particle Hilbert space $\mathcal{H}$ with positive-frequency modes $u_{\mathbf{k}}(t,\mathbf{x})$ —relative to the inertial time t— by

$$u_{\mathbf{k}}(t,\mathbf{x}) \equiv \frac{1}{4\pi^{3/2}|\mathbf{k}|^{1/2}} e^{-i|\mathbf{k}|t} e^{i\mathbf{k}\cdot\mathbf{x}}, \qquad (5.41)$$

where $\mathbf{k} \in \mathbb{R}^3$ and $u_{\mathbf{k}}(t,\mathbf{x})$ form a complete basis for $\mathcal{H}$. If the field is at thermal equilibrium at temperature T —it is in a KMS state— ν_B will be [38]

$$\nu_B = \exp\left\{-2\left\langle KEf_B\middle|\coth\left(\frac{\beta\hat{h}}{2}\right)KEf_B\right\rangle\right\}, \qquad (5.42)$$

where β is the inverse of the temperature T and $\hat{h} : \mathcal{H} \to \mathcal{H}$ is the one-particle Hamiltonian, which satisfies

$$H_\phi = 1 \oplus \hat{h} \oplus (\hat{h} \oplus \hat{h}) \oplus ..., \qquad (5.43)$$

also, $\hat{h} = i\partial_t$ and $K : S^{\mathbb{C}} \to \mathcal{H}$ [13]. Additionally, we can decompose KEf_B in terms of $u_{\mathbf{k}}$, since it comprehends a complete basis for the Hilbert space, thus,

$$|KEf_B\rangle = \int d^3\mathbf{k}\, \langle u_{\mathbf{k}}|Ef_B\rangle\, |u_{\mathbf{k}}\rangle\,. \qquad (5.44)$$

In this way, we will have

$$\left\langle KEf_B\middle|\coth\left(\frac{\beta\hat{h}}{2}\right)KEf_B\right\rangle = \int d^3\mathbf{k}\coth\left(\frac{\beta|\mathbf{k}|}{2}\right)|\,\langle u_{\mathbf{k}}|Ef_B\rangle\,|^2, \qquad (5.45)$$

where, we can use the properties of E, described in chapter 3, to write

$$\begin{aligned}
\langle u_{\mathbf{k}}|Ef_B\rangle &= i\int \epsilon_{\mathbb{M}}\, \overline{u_{\mathbf{k}}} f_B && (5.46)\\
&= i\int \epsilon_{\mathbb{M}}\, \frac{e^{i|\mathbf{k}|t}e^{-ik_x x}}{2^{3/2}\pi\sqrt{2\pi}\mathbf{k}^{1/2}}\left[\epsilon_B c_B(t)\delta^3(\mathbf{x}-L\hat{\mathbf{x}})\right] && (5.47)\\
&= \frac{i}{2^{3/2}\pi\mathbf{k}^{1/2}\sqrt{2\pi}}\int d^3\,\mathbf{x}e^{-ik_x x}\epsilon_B\delta^3(\mathbf{x}-L\hat{\mathbf{x}})\int_{\mathbb{R}} dt\, e^{i|\mathbf{k}|t}c_B(t) && (5.48)\\
&= \frac{i\epsilon_B}{2^{3/2}\pi\mathbf{k}^{1/2}}e^{-ik_x L}\tilde{c}_B(|\mathbf{k}|), && (5.49)
\end{aligned}$$

with

$$\tilde{c}_B(|\mathbf{k}|) = \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} dt\, e^{i|\mathbf{k}|t}c_B(t). \tag{5.50}$$

Therefore, we will get

$$\left\langle KEf_B\middle|\coth\left(\frac{\beta\hat{h}}{2}\right)KEf_B\right\rangle = \int d^3\mathbf{k}\coth\left(\frac{\beta|\mathbf{k}|}{2}\right)\frac{\epsilon_B^2|\tilde{c}_B(|\mathbf{k}|^2)|}{2^3\pi^2|\mathbf{k}|}, \tag{5.51}$$

such that

$$\nu_B = \exp\left\{-\frac{\epsilon_B^2}{2^2\pi^2}\int d^3\mathbf{k}\coth\left(\frac{\beta|\mathbf{k}|}{2}\right)\frac{|\tilde{c}_B|^2}{|\mathbf{k}|}\right\}. \tag{5.52}$$

If we use $d^3\mathbf{k} = 4\pi k^2 dk$, where $k \equiv |\mathbf{k}|$, the above equation will become [13]

$$\nu_B(T) = \exp\left\{-\frac{2\epsilon_B^2}{\pi}\int_0^\infty dk\, k\coth\left(\frac{k}{2T}\right)|\tilde{c}_B(k)|^2\right\} \tag{5.53}$$

The results in Eqs. (5.40) and (5.53) can be used to analyse the dependence of the channel capacity $C(\mathcal{E})$ with Alice and Bob's coupling constants ϵ_A and ϵ_B, respectively.

This dependence can be seen in Fig. 5.5, where it was considered a quantum field that is initially in the inertial vacuum state, i.e., $\beta \to \infty$, so that ν_B resumes to [13]

$$\nu_B = \exp\{-2\langle KEf_B|KEf_B\rangle\}. \tag{5.54}$$

Note that, for large values of ϵ_A and lower values of ϵ_B the channel capacity increases almost to its maximum value. Thus, the process of transmitting information will be more efficient if the interaction of Alice's qubit with the field is strong enough, while the interaction of Bob's qubit is acceptably weak, so that he receives the information but does not lose it by decoherence —for further details see [13]. Therefore, we can conclude that, for large values of ϵ_A and small values of ϵ_B the channel capacity increases which also boost the efficiency of the heat engine in Bob's laboratory.

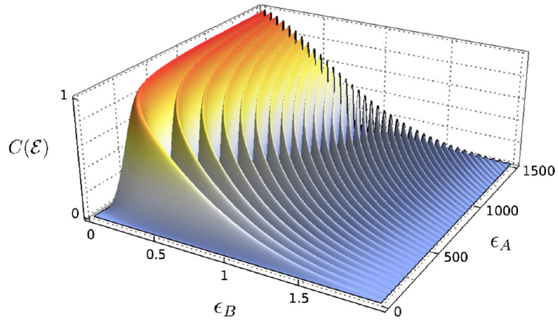

Figure 5.5: Dependence of the Channel Capacity with the coupling constants ϵ_A and ϵ_B. (Source: I. B. Barcellos and A. G. Landulfo, *"Relativistic quantum communication: energy cost and channel capacities"*)

Another interesting fact is that, since we are considering a massless field, the information imprinted on it will travel on light's world-line and Alice's qubit interaction with the field will create gap region where the encoded message can be measured, as it is illustrated by Fig. 5.6.

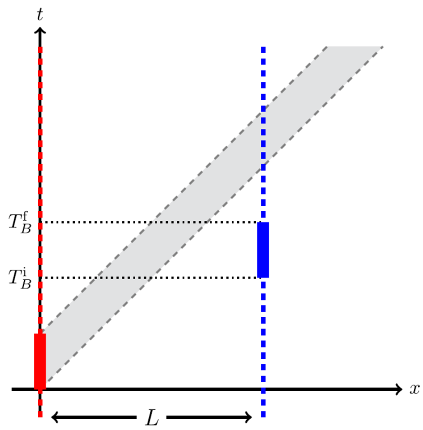

Figure 5.6: Causal relation between the emission and detection events. (Source: I. B. Barcellos and A. G. Landulfo, *"Relativistic quantum communication: energy cost and channel capacities"*)

From the above figure we can see the regions where the qubits interacts with the field, the red rectangle represents Alice's and the blue one represents Bob's. The gray region, whose size depends of Alice's interaction time, describes the region which contains the encoded message, if

Bob wants to get the whole message he needs to interacts with the field exactly in this region. In other words, the emission and detection events must be causal related so that the channel capacity will not be zero and, consequently, Bob can use the Landauer principle to run his heat engine and produce work. These events are spacelike when $T_B^i < L - \Delta t$, where $\Delta t = T_B^f - T_B^i = T_A^f - T_A^i$, and timelike when $T_B^i > L + \Delta t$ [13], thus, in order to produce the maximum amount of work Bob must switch on his qubit interaction in $T_B^i = L$ and switch it off in $T_B^f = L + \Delta t$.

Chapter 6

Conclusions and Future Perspectives

Throughout this monograph, we have studied several things. First we presented the information measures and the basic procedure of the communication between Alice and Bob, so that we could introduce, and demonstrate, both Shannon theorems.

We saw that the process of compressing information can result in a bad efficiency of the channel, i.e., some of the information can be lost such that Bob would not understand the message, and to solve that, Shannon proposed the compression theorem, which gives us a lower bound on the compression rate, so the information won't be lost. Also, we saw that, the channel capacity theorem determines an upper bound on the reliable information which can be transmitted per use of the channel, i.e., the maximum amount of reliable information that Bob could get from Alice.

After that, we introduced the postulates of general relativity, such that we could present the spacetime structure $(\mathbb{M}, g_{\mu\nu})$, which were used to formulate a quantum field theory of the Klein-Gordon scalar field ϕ. Later, we described the quantum communication channel, $\mathcal{E}$, which were used to calculate its capacity and the energy cost for transmitting information on it. We also concluded that, when Alice and Bob were causally disconnected, the channel capacity will vanish, which is expected since, in this situation, they wouldn't be able to communicate with each other. Also, we separated the total energy variation of the system into three contributions. Where the first one is due to the change of the spacetime metric when we go from I^- to $\mathfrak{h}$. The second one is due to the work necessary to switch on or off the qubit interaction with the field, and the last one is the contribution of the communication process itself.

The energy contribution due to the process of communication will be zero when we choose an initial state for Bob's qubit which maximizes the channel capacity. Thus, we concluded that, after the quantum computation system is settled, i.e., after the qubits are created for some purpose, Alice and Bob won't expend any extra energy by using it to convey information. But the Landauer's principle says that, any irreversible manipulation of information, such as the erasure of information, must result in a corresponding production of heat.

Therefore, if we consider that the quantum communication system is set up and Bob has access to two apparatus at the same temperature T_0. He could use the heat, which were generated by erasing the information, to increase the temperature of one system to $T_1 > T_0$. Thus, since he has two systems at different temperatures, he would be able to produce work $W > 0$. We could think that he would be violating the second law of thermodynamics by doing this process, but we saw that there is an energy cost of $k_B T \ln 2$ to erase each bit of information. Thus, if the heat engine absorbs an amount of heat equal to Q, described in Eq. (5.7), the maximum work produced by it must be equal to ηQ and, since $\eta < 1$, Bob will not produce more energy than he spend to erase the information. In this way, we showed that the second law of thermodynamics will not be violated.

Appendix A

Manifolds

Manifolds are a fundamental concept in mathematics and are used extensively in physics, including general relativity. A manifold is a mathematical space that locally looks like Euclidean space but globally may have a very different structure. In other words, a manifold is a space that can be described by a set of coordinates, just like Euclidean space, but those coordinates can only be used locally, and different regions of the manifold may require different sets of coordinates.

In general relativity, manifolds play a crucial role in describing the geometry of spacetime. According to Einstein's theory of general relativity, the presence of matter and energy warps the geometry of spacetime, and this curvature is described by a mathematical object called the metric tensor. The metric tensor is defined on the manifold that represents spacetime, and it determines how distances and angles are measured on the manifold.

Without the concept of manifolds, it would be impossible to describe the curvature of spacetime in a mathematically rigorous way. Differential manifolds provides the framework for the precise mathematical description of spacetime geometry that is needed for the theory of general relativity. As mentioned in Chapter 3, since we write the laws of physics in terms of derivatives, it is fundamental that we describe the spacetime structure as a differential manifold. Some examples of manifolds includes the surface of a sphere and the surface of a torus.

To formalize the concept of manifolds first let us introduce the notion of continuity of a map as well as open sets. Consider the map $f : \psi_m \to \psi_n$, it takes an m-tuple $(x_1, x_2, ..., x_m)$ to an n-tuple

$(y_1, y_2, ..., y_n)$, such that [23, 25]

$$\begin{aligned}
y_1 &= f_1(x_1, x_2, ..., x_m) \\
y_2 &= f_2(x_1, x_2, ..., x_m) \\
&\vdots \\
y_n &= f_n(x_1, x_2, ..., x_m).
\end{aligned} \tag{A.1}$$

If the nth derivative of these functions exists and is continuous we refer to the map f as C^n. Thus a C^∞ map is continuous and infinitely differentiable. These kind of maps are usually called *smooth* [25].

An open set is a subset of a topological space that does not contain any of its boundary points. More formally, a set $\mathbb{O}$ in a topological space $\mathbb{M}$ is said to be open if for every point x in $\mathbb{O}$, there exists an open neighborhood of x, i.e., a set that contains an open ball centered at x, that is contained entirely within $\mathbb{O}$. To put it simply, a set is open if it contains points that are not on its boundary. For example, the open interval $(0, 1)$ on the real line is an open set, because it does not contain its endpoints 0 and 1. Some key properties of open sets are:

i. An empty set $\emptyset$ is always open;

ii. The union of any collection of open sets is open;

iii. The intersection of a finite collection of open sets is open;

iv. If $\mathbb{U}$ and $\mathbb{V}$ are open sets, then their intersection $\mathbb{U} \cap \mathbb{V}$ is also open;

v. The complement of a closed set is always open, and vice versa.

Now we can define an n-dimensional, C^∞, manifold $\mathbb{M}$ as a set made up of a collection of subsets $\{\mathbb{O}_n\}$ which are "smoothly" connected and satisfies the following properties [23]:

1. Each point $x \in \mathbb{M}$ lies in at least one open set $\mathbb{O}_n$;
2. If $\mathbb{E}_n$ is an open set of the real space R^n, for each n there is a one-to-one, onto, map $\psi_n : \mathbb{O}_n \to \mathbb{E}_n$;

3. If two sets $\mathbb{O}_i$ and $\mathbb{O}_j$ overlap, i.e., $\mathbb{O}_i \cap \mathbb{O}_j = \emptyset$, we can define a map ψ_{ij} which takes us from points in $\psi_i[\mathbb{O}_i \cap \mathbb{O}_j] \subset \mathbb{E}_i \subset \mathbb{R}^n$ to points in $\psi_j[\mathbb{O}_i \cap \mathbb{O}_j] \subset \mathbb{E}_j \subset \mathbb{R}^n$.

The collection $\{\mathbb{O}_n\}$ of subsets is said to be "smoothly" connected because we demand that the maps ψ_{ij} are C^∞ [23]. Additionally, in physics, maps ψ_n are usually called *coordinate systems* and ψ_{ij} are the *coordinate transformations*.

Bibliography

[1] C. E. Shannon, “A mathematical theory of communication”, The Bell system technical journal **27**, 379–423 (1948).

[2] V. Vedral et al., *Introduction to quantum information science* (Oxford University Press on Demand, 2006).

[3] M. M. Wilde, “From classical to quantum shannon theory”, arXiv preprint arXiv:1106.1445 (2011).

[4] C. Adami, “Information theory in molecular biology”, Physics of Life Reviews **1**, 3–22 (2004).

[5] M. A. Nielsen and I. Chuang, *Quantum computation and quantum information*, 2002.

[6] E. Tjoa and K. Gallock-Yoshimura, “Channel capacity of relativistic quantum communication with rapid interaction”, Physical Review D **105**, 085011 (2022).

[7] C. Dappiaggi, V. Moretti, and N. Pinamonti, “Cosmological horizons and reconstruction of quantum field theories”, Communications in Mathematical Physics (2009).

[8] A. Einstein et al., “On the electrodynamics of moving bodies”, Annalen der physik (1905).

[9] A. Einstein, “The field equations of gravitation”, Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys.) **1915**, 844–847 (1915).

[10] J. B. Hartle, *Gravity: an introduction to einstein’s general relativity*, 2003.

[11] A. G. Landulfo, “Nonperturbative approach to relativistic quantum communication channels”, Physical Review D (2016).

[12] R. M. Wald, *Quantum field theory in curved spacetime and black hole thermodynamics* (University of Chicago press, 1994).

[13] I. B. Barcellos and A. G. Landulfo, “Relativistic quantum communication: energy cost and channel capacities”, Physical Review D (2021).

[14] R. Landauer, “Irreversibility and heat generation in the computing process”, IBM journal of research and development **5**, 183–191 (1961).

[15] H. B. Callen, *Thermodynamics and an introduction to thermostatistics*, 1998.

[16] M. B. Plenio and V. Vitelli, “The physics of forgetting: landauer’s erasure principle and information theory”, Contemporary Physics **42**, 25–60 (2001).

[17] R. Bousso, H. Casini, Z. Fisher, and J. Maldacena, “Entropy on a null surface for interacting quantum field theories and the bousso bound”, Physical Review D **91**, 084030 (2015).

[18] P. Pal and A. Jayannavar, “Maxwell’s demon, szilard engine and landauer principle”, Resonance **26**, 443–462 (2021).

[19] H. Casini, “Relative entropy and the bekenstein bound”, Classical and Quantum Gravity **25**, 205021 (2008).

[20] H. Pishro-Nik, “Introduction to probability, statistics, and random processes”, (2016).

[21] J. C. Maxwell, “Viii. a dynamical theory of the electromagnetic field”, Philosophical Transactions of the Royal Society of London.

[22] R. Resnick, “Introduction to special relativity”, (1971).

[23] R. M. Wald, *General relativity* (University of Chicago press, 2010).

[24] L. D. Landau and E. M. Lifshitz, *Course of theoretical physics* (Elsevier, 2013).

[25] S. M. Carroll, *Spacetime and geometry* (Cambridge University Press, 2019).

[26] C. J. Fewster, “Lectures on quantum field theory in curved spacetime”, (2008).

[27] M. Nakahara, *Geometry, topology and physics* (CRC press, 2018).

[28] B. Schutz, *A first course in general relativity* (Cambridge university press, 2009).

[29] J. J. Sakurai, *Advanced quantum mechanics* (Pearson Education India, 2006).

[30] E. Poisson, *A relativist's toolkit: the mathematics of black-hole mechanics* (Cambridge university press, 2004).

[31] E. Tjoa and K. Gallock-Yoshimura, “Channel capacity of relativistic quantum communication with rapid interaction”, Physical Review D **105**, 085011 (2022).

[32] A. Bérut, A. Arakelyan, A. Petrosyan, S. Ciliberto, R. Dillenschneider, and E. Lutz, “Experimental verification of landauer's principle linking information and thermodynamics”, Nature **483**, 187–189 (2012).

[33] C. H. Bennett, “Notes on landauer's principle, reversible computation, and maxwell's demon”, Studies In History and Philosophy of Science Part B: Studies In History and Philosophy of Modern Physics **34**, 501–510 (2003).

[34] H. M. Nussenzveig, *Curso de física básica: fluidos, oscilaçõs e ondas, calor*, Vol. 2 (Editora Blucher, 2018).

[35] R. Clausius, “I. on the moving force of heat, and the laws regarding the nature of heat itself which are deducible therefrom”, The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science **2**, 1–21 (1851).

[36] J. M. Calm, “Heat recovery in air conditioning and refrigeration”, Energy Conservation Measures, 159–169 (1984).

[37] S. A. Fulling et al., *Aspects of quantum field theory in curved spacetime*, 17 (Cambridge university press, 1989).

[38] B. S. Kay, “A uniqueness result for quasi-free kms states”, Helvetica Physica Acta **58**, 1017–1029 (1985).