

ALMA MATER STUDIORUM · UNIVERSITÀ DI BOLOGNA

---

Scuola di Scienze  
Dipartimento di Fisica e Astronomia  
Corso di Laurea in Fisica

Quantum Information and Quantum  
Communication Theory in relation to black  
hole idealized mechanical models

Relatore:  
Prof. Francesco Ravanini

Candidato:  
Leonardo Bersigotti

Anno Accademico 2021/2022



# Abstract

Nowadays, the concept of information has become crucial in physics. Moreover, because our best theory to make predictions about the universe is quantum mechanics, it becomes crucial to develop a quantum version of information theory. This centrality is confirmed by the fact that black holes do have entropy. In this work, elements of quantum information and quantum communication theory are given and some are illustrated by referring to highly idealized quantum models of black hole mechanics. In the first chapter all the quantum-mechanical tools for quantum communication and information theory are given. Later, we discuss about quantum information theory and we arrive at the Bekenstein's bound on amount of information enclosed in a spatial region. We address the problem by using a quantum idealized black hole mechanics model assisted by thermodynamics. In the last chapter, we deal with the problem of finding an achievable rate for quantum communication and we use again an idealized quantum model of black hole in order to illustrate elements of the theory. Lastly, a short summary of black hole physics is given in the appendix.



# Sommario

Oggigiorno il concetto di informazione è diventato cruciale in fisica, pertanto, siccome la migliore teoria che abbiamo per compiere predizioni riguardo l'universo è la meccanica quantistica, assume una particolare importanza lo sviluppo di una versione quantistica della teoria dell'informazione. Questa centralità è confermata dal fatto che i buchi neri hanno entropia. Per questo motivo, in questo lavoro sono presentati elementi di teoria dell'informazione quantistica e della comunicazione quantistica e alcuni sono illustrati riferendosi a modelli quantistici altamente idealizzati della meccanica di buco nero. In particolare, nel primo capitolo sono forniti tutti gli strumenti quanto-meccanici per la teoria dell'informazione e della comunicazione quantistica. Successivamente, viene affrontata la teoria dell'informazione quantistica e viene trovato il limite di Bekenstein alla quantità di informazione chiudibile entro una qualunque regione spaziale. Tale questione viene trattata utilizzando un modello quantistico idealizzato della meccanica di buco nero supportato dalla termodinamica. Nell'ultimo capitolo, viene esaminato il problema di trovare un tasso raggiungibile per la comunicazione quantistica facendo nuovamente uso di un modello quantistico idealizzato di un buco nero, al fine di illustrare elementi della teoria. Infine, un breve sommario della fisica dei buchi neri è fornito in appendice.



# Contents

<b>Introduction</b>	<b>13</b>
<b>1 Foundations of Quantum theory</b>	<b>15</b>
1.1 Closed Quantum Systems	15
1.2 Open Quantum systems	17
1.2.1 The density operator	17
1.2.2 Projective measurements and tracing operations	18
1.2.3 Qubits	20
1.2.4 Schmidt decomposition and Maximal Entanglement	22
1.2.5 Fidelity of Quantum States	23
1.2.6 Generalized measurements and POVM	24
1.2.7 Quantum channels	26
1.2.8 Axioms revisited	28
1.3 Entanglement-assisted communication	28
1.3.1 Hidden Quantum Information and <i>LOCC</i>	28
1.3.2 Dense Coding and Quantum Teleportation	30
<b>2 Dawn of Quantum Information Theory</b>	<b>33</b>
2.1 Shannon Theory	33
2.1.1 Noisy channel coding	35
2.2 Von Neumann Entropy	38
2.2.1 Entropy and thermodynamics	39
2.2.2 Bekenstein's entropy bound	40
<b>3 Quantum Channel Capacities and Decoupling</b>	<b>43</b>
3.1 Quantum Capacity Theorem and Decoupling approach	44
3.2 The Decoupling Inequality	45
3.3 Black holes as mirrors	47
<b>Conclusions and Overviews</b>	<b>51</b>
<b>A Summary of black holes physics</b>	<b>53</b>
A.1 Black hole thermodynamics	53
A.1.1 Bekenstein and spherical bound for entropy of matter systems	55

---

A.1.2	Hawking radiation and evaporation . . . . .	57
A.1.3	Quantum states of a black hole and thermalization . . . . .	59
A.2	The problem of unitarity and black hole complementarity . . . . .	60



# Acknowledgements

Words cannot express my gratitude to my supervisor Mr. Francesco Ravanini for his invaluable patience and feedback. This thesis would not have been possible without the help and encouragement he gave me.

I am also grateful to my friends who have been close to me and offered deep insight into the study.

Lastly, I could not have undertaken this journey without my family, especially my parents, my sister and my girlfriend. Their belief in me has kept my spirits high during this process and they pushed me to finally realize this work.



# Introduction

When we hear about physics, we might not think immediately about information. Whereas, at present day, our best theory of the microscopic world- quantum mechanics, that describes how atoms and particles interact through the forces of nature and makes incredibly precise experimental predictions- is based on the foundations of probability and entropy. This makes it an *inferential* theory. That is, rather than being a description of the behavior of the universe, this theory describes how observers can make optimal predictions about the universe. In such a scenario, information plays a critical role. To be more specific, with *information* we denote the number of yes/no questions we need to get answered to fully specify a physical system. For this reason, we use to collect the answers of these questions in strings of *bits*, each of which can assume value 0 or 1. What is more, there are little hints, such as the fact that black holes have entropy, that continue to suggest that information is fundamental to physics in general. The upshot is that information is physical and strongly related to the state of a physical system, so that the evolution of a physical process has to tell us about information. In fact, as states evolves, the information they own has to be transformed without being lost. Thus, information, and therefore every bit, is indestructible. That is the most important law of physics. Correspondingly, one might also reproduce (through another transformation operation) at one point either exactly or approximately a certain piece of information (*i.e.*, a message) with meaning at another point. The fact that the message has a meaning denotes that it refer to or is correlated according to some system (*i.e.*, the code-system) with certain physical or conceptual entities [1]. This is the problem of communication which introduces us to some interesting issues, such as how many bits do occur in order to send a message and how many so that this message could be correctable after being somehow corrupted the transfer process. Both this issues are addressed by the two *source coding* and *noisy channel coding* Shannon's theorems. In such a picture, there must be included the fact that essentially the universe is quantum mechanical. Therefore, the classical ideas about information would need revision under the new physics and has to considers features like quantum states non distinguishability and entanglement entropy. Moreover, since acquiring information causes disturbance by quantum theory statements, we cannot make a perfect copy of a quantum state [2] since we could measure an observable of the copy without disturbing the original, defeating the principle of disturbance.

In a quantum context, in order to quantify how much we do not know about the informative content of a physical state, we use Von Neumann entropy, therefore quantum information theory is inevitably related with but also illuminates thermodynamics. Of the utmost importance is the case of black holes, where the laws of mechanics are surprisingly similar to the thermodynamical ones. In particular, the never-decreasing-area law is the same as the second law for entropy. Moreover, Bekenstein proposed that this link between the horizon area and entropy was not only formal [3]. Therefore, we find that physics, through thermodynamics, imposes a limit on information that can be encoded in a spatial region (that might also not be occupied by a black hole). That is the Bekenstein's entropy bound which can also be obtained from quantum information tools and statements.

Entropy has also an important role in quantum communication theory. In fact, the highest achievable quantity of information that can be shipped through a quantum communication channel by using a single *qubit* (*i.e.*, the quantum mechanical generalization of a bit) which allows to obtain an errorless transmission, is the *channel capacity*. Entropy enters in this topic because the capacity is related to the *conditional entropy* between the message source and receiver systems. A formula for such a capacity can be found and demonstrated by using a simple principle that assume that the transmitted data will with high probability be decoupled from the channel's environment. This principle constitutes a special tool that can be applied and used also for simplified real representations of quantum channels with noise. By this way, the information retrieval from evaporating black holes can be studied and it is such an fascinating situation to inspect in order to see all the quantum information theory power applied.

Therefore, in this work many central points of both classical and quantum information theory are discussed. We start by doing a brief review of quantum mechanics, inspecting how the axioms of the theory change under the transition from closed to open quantum systems [4][5]. The difference is essentially in how much we know about the system resulting from considering or not its environment. Then, we focus ourselves mainly on open quantum systems as they are more realistic. Qubits and their geometrical interpretation through Bloch sphere are widely discussed, while particular attention is given to the reformulation of the state evolution concept in open systems which becomes a quantum channel that allows states to decohere and switch from pure into mixed [6]. From this, we put in evidence the basis of entanglement-assisted communication, by explaining how entanglement enhances and redefines protocols used for communication [7]. At this point, we have all we need in our hands in order to study information and how to manipulate it. In the second chapter, giving a brief introduction of classical Shannon theory, we soon come to its quantum definition and so we introduce the Von Neumann entropy [8]. This inevitably leads to thermodynamics, moreover links between this and information theory are discussed talking about the Bekenstein bound on information. In the last chapter, we inspect quantum channel capacity and the so called *decoupling ap-*

---

*proach* reaching the final application using a black hole as a quantum channel with high erasure probability [9]. More in detail, we studied information retrieval from a black hole that has already radiated more than half of its initial number of degrees of freedom. To conclude, from black holes physics we have that if bits are forever, information get lost when they fall into a black hole. This was Hawking's central point in 1976 when he created the information paradox [10]. In order to briefly answer to this open question and to furnish a brief review about black holes dynamics, an Appendix has been added to this work. It is important to underline that the purpose for this appendix is not to furnish accurate relativistic features of black hole physics, which are beyond the possibilities of such a work. Instead, it aims to focus attention on the quantum properties of black holes, and try to give a review of the developments originated from the reformulation of black hole's mechanics in terms of thermodynamics.



# Chapter 1

## Foundations of Quantum theory

Now we need to be more specific about the mathematical description of quantum information and its correlations with quantum system evolution. A brief introduction to the foundations of quantum mechanics is provided. Indeed, we introduce the fundamental points of the theory (state, observable, measurement, dynamics and system composition) starting from considering a *closed* system and later an *open* system. The motivation for studying closed and open systems is that all realistic systems are open, as it is impossible to perfectly isolate them from their environment, but, to understand the behavior of an open system, we should regard its combination with its environment as a closed system. Then, we wonder how the system would behave without considering its environment. Closed quantum systems are easier to study and for that we start from this.

### 1.1 Closed Quantum Systems

Quantum theory is the mathematical model that represents the physical world. According to this, each physical system is associated with a vector space over the complex numbers called *Hilbert space*  $\mathcal{H}$ , equipped with an inner product which defines a norm which turns  $\mathcal{H}$  into a complete metric space. A *pure state* of a quantum system is an equivalence class of vectors in  $\mathcal{H}$  (*i.e.*, a *ray*) that differ by multiplication by a nonzero complex scalar. That holds in *perfectly isolated* quantum systems (*i.e.*, Closed Quantum Systems, later denoted as CQS). Using Dirac's bra-ket notation, vectors in  $\mathcal{H}$  are denoted with  $|\psi\rangle$  (namely, we say *ket psi*). By convention, we can choose a representative of the equivalence class that has unit norm. Therefore, *pure* states correspond to unit vectors in  $\mathcal{H}$  and the set of all pure states corresponds to the unit sphere in the Hilbert space.

The general principle of *quantum superposition* applies to the states: whenever the system is definitely in one state, we can consider it as being partly in each of two or more other states [11]. Hence, if a basis is chosen in a finite-dimensional Hilbert

space, any ket  $|\psi\rangle$  can be expressed as a linear combination of basis element  $|k_i\rangle$  (assuming them to be orthonormal without loss of generality) as

$$|\psi\rangle = \sum_i c_i |k_i\rangle \quad (1.1)$$

where  $c_i$  are complex numbers. As we consider normalized states  $|\psi\rangle$ , we must have  $\sum_i |c_i|^2 = 1$ . The meaning of normalization could be expressed in terms of quantum-mechanical probability, as the total probability of the system has to be  $\langle\psi|\psi\rangle = 1$ . For this reason, we have that  $e^{i\theta}|\psi\rangle = |\psi\rangle$ . However, the members of the above equation are not interchangeable in combinations of different states (see Open Quantum Systems (OQS)), therefore we say that global phase factors are unphysical, but relative phase factors are physical and relevant.

Every state encodes information about physical quantities belonging to the system that can be in principle measured; these are the *observables*. In quantum mechanics, an observable is a self-adjoint operator  $\mathbf{A}$  in  $\mathcal{H}$ . For this, its eigenstates form a complete orthonormal basis of the Hilbert space and can therefore be identified with the  $|k_i\rangle$  (with eigenvalues  $a_i$ ). We have

$$\mathbf{A} = \sum_i a_i \mathbf{E}_i \quad (1.2)$$

where the  $\mathbf{E}_i$  are the orthogonal projection onto the space of eigenvectors corresponding to eigenvalues  $a_i$ 's ([4], section 2.1).

The process in which information about the state of a physical system is acquired by an observer is called *measurement*. The measurement of an observable  $\mathbf{A}$  prepares an eigenstate of  $\mathbf{A}$ , and the observer learns the value of the corresponding eigenvalue. So, measurements in closed quantum systems are *orthogonal projection*. For example, using orthonormality of the  $|k_i\rangle$  and (1.1), the probability that the outcome  $a_i$  is obtained from a measurement of  $|\psi\rangle$ , is

$$Prob(a_i) = \langle\psi|\mathbf{E}_i|\psi\rangle = |c_i|^2 \quad (1.3)$$

In closed quantum systems, the *evolution of states over time* is described by a unitary operator  $\mathbf{U}(t', t)$ . It transforms the initial state at time  $t$  into a final state at time  $t'$ . Infinitesimal time evolution is governed by the *Schrödinger equation*

$$\frac{d}{dt} |\psi\rangle = -i\mathbf{H}(t) |\psi\rangle$$

that can be expressed to first order as:  $|\psi(t + dt)\rangle = (\mathbf{I} - i\mathbf{H}(t)dt) |\psi\rangle$ .

Therefore, the operator  $\mathbf{I} - i\mathbf{H}(t)dt$  is unitary and the evolution governed by the *Schrödinger equation* over a finite interval is also unitary.

If we have a *composite system* AB and the Hilbert space of the system A is  $\mathcal{H}_A$ , while the Hilbert space of the system B is  $\mathcal{H}_B$ , then the Hilbert space of the composite system AB is the tensor product

$$\mathcal{H}_A \otimes \mathcal{H}_B$$



Fixed a basis  $\{|i\rangle_A\}$  for  $\mathcal{H}_A$  and a basis  $\{|j\rangle_B\}$  for  $\mathcal{H}_B$ , the most general state in  $\mathcal{H}_A \otimes \mathcal{H}_B$  has the form

$$|\psi\rangle_{AB} = \sum_{i,j} c_{ij} |i\rangle_A \otimes |j\rangle_B \quad (1.4)$$

We say it is *separable* if we can rightly rewrite

$$|\psi\rangle_{AB} = \sum_{i,j} c_i^A c_j^B |i\rangle_A \otimes |j\rangle_B = \sum_i c_i^A |i\rangle_A \otimes \sum_j c_j^B |j\rangle_B = |\psi\rangle_A \otimes |\phi\rangle_B \quad (1.5)$$

If a state cannot be expressed as a direct product of pure states in  $\mathcal{H}_A$  and  $\mathcal{H}_B$ , we say it is *entangled*.

To resume, all the fundamental assumptions of quantum mechanics for closed quantum systems are:

- (1) A *state* is a ray in  $\mathcal{H}$ ;
- (2) An *observable* is a self-adjoint operator on  $\mathcal{H}$ ;
- (3) A *measurement* is an orthogonal projection;
- (4) *Time evolution* is unitary;
- (5) A *composite system*  $AB$  is described by the tensor product  $\mathcal{H}_A \otimes \mathcal{H}_B$ .

## 1.2 Open Quantum systems

### 1.2.1 The density operator

If we confine our attention only on a part of larger quantum system, then we are considering an OQS. Considering the closed system  $AB$ , that is a quantum system  $A$  interacting with its environment  $B$ , then the information of the system  $A$  is lost in the environment. Therefore,  $AB$  is a CQS described by a pure state, but  $A$  is not. In this case, assumptions (1-4) need not to be satisfied ([5]: section 2.3.1). We have partial ignorance of the preparation of  $A$ . In particular, the state of  $A$  could be a weighted superposition of different pure states, up to obtaining a statistical *ensemble* of different state vectors. In this case we call this a *mixed state* (remember that relative phase factors are physical, hence  $|\phi\rangle + |\psi\rangle \neq |\phi\rangle + e^{i\theta} |\psi\rangle$ ). It occurs a new mathematical description for such states. This is encoded in the *density matrix* (or *density operator*)  $\rho$ . It is a self-adjoint, positive matrix with unit trace (because we chose normalized states). Every density matrix can be expressed in the basis in which it is diagonal as a sum of pure states

$$\rho = \sum_i w_i |a_i\rangle \langle a_i| \quad (1.6)$$

where the  $w_i$  are positive-valued probabilities and they sum up to 1, while the vectors  $a_i$  are unit vectors of  $\mathcal{H}$ . The upshot is that we can interpret  $\rho$  representing an ensemble where pure states  $|a_i\rangle$  are prepared with probability  $w_i$ .

Another case in which mixed states arise is entanglement. As in (1.5), in an entangled state  $AB$  is not possible to find a coefficient matrix so that  $c_{ij} = c_i^A c_j^B$ . So, it is impossible to describe the subsystem of an entangled system as a pure state (in the form (1.1)). Quantum entanglement prevents the complete knowledge about the subsystem by making some or all of the coefficients of the superposition (1.1) inaccessible.

However, density matrices are not only involved in the description of mixed states. For example, the partial ignorance of a state preparation, does not exclude that there could theoretically be another person who knows the full history of  $A$ , and therefore describe the state of the subsystem as a pure state. When  $|\psi\rangle_A$  of  $A$  is pure, then the density matrix is the projection onto the one-dimensional space spanned by  $|\psi\rangle_A$

$$\rho_A = |\psi\rangle_A \langle\psi|_A \quad (1.7)$$

Therefore, there is only one term in the sum (1.6). Hence, the density operator is said *pure* and

$$\rho^2 = \rho \quad (1.8)$$

Follows that mixed quantum state has two or more terms in the sum (1.6) and in this case (1.8) does not hold. Consequently, we say that  $\rho$  is an incoherent mixture of the states  $|a_i\rangle$ . In fact, the relative phases of the  $|a_i\rangle$  are experimentally inaccessible. For what said, the entanglement between  $A$  and  $B$  destroys the coherence of a superposition of states of  $A$ , so that some of the phases in the superposition become inaccessible if we look at  $A$  alone.

## 1.2.2 Projective measurements and tracing operations

Measurement in quantum mechanics have particularly simple rules in terms of density matrices. We start by defining  $O$  as an alphabet set of all possible outcomes of the measurements. A projection-valued measure (PVM) is the simplest quantum measurement, defined as a complete set of positive semi-definite Hermitian matrices  $\{\mathbf{E}_i, i \in O\}$ , describing elementary projectors, so that  $\mathbf{E}_a \mathbf{E}_b = \delta_{ab} \mathbf{E}_a$ ; the ones of (1.2) [12]. Also in (1.2) we saw the probability of obtaining outcome  $a_i$  for a measurement performed on a pure state. For a mixed state described by  $\rho_A$ , the probability distribution over the outcomes can be computed from the density operator. Therefore, we can rewrite (1.3) as

$$\text{Prob}(a_i) = \text{tr}(\mathbf{E}_i \rho_A) \quad (1.9)$$

The average of the eigenvalues of an observable  $\mathbf{A}$ , weighted by the above probabilities, is the expectation value of that observable

$$\langle \mathbf{A} \rangle = \text{tr}(\mathbf{A} \rho_A) \quad (1.10)$$

where we see that it is possible to express the expectation value of an observable as the result of a trace operation. Actually, considering a bipartite quantum system in  $\mathcal{H}_A \otimes \mathcal{H}_B$ , the equation (1.10) provides only a compact way to characterize a measurement performed on one of the subsystems, we choose  $A$ . An orthonormal basis for the composite space can be found in  $\{|i\rangle_A \otimes |j\rangle_B\}$ . Therefore one can rightly rewrite the projectors  $\mathbf{E}_i$  as

$$|i\rangle \langle i| \otimes \mathbf{I}_B \quad (1.11)$$

where  $|i\rangle \langle i|$  are the orthogonal projector onto the one-dimensional space spanned by  $|i\rangle$  and  $\mathbf{I}_B$  is the identity operator acting on  $B$ . Consequently, an observable  $\mathbf{A}$  acting on  $A$  is more precisely expressed by considering the action of a suitable composite operator on the larger system  $AB$

$$\mathbf{A} \otimes \mathbf{I}_B \quad (1.12)$$

A normalized pure state  $|\psi\rangle$  of  $AB$  can be expressed as (1.4), where now  $\sum_{i,j} |c_{ij}|^2 = 1$  is inevitably. The expectation value of  $\mathbf{A}$  in this state is

$$\begin{aligned} \langle \mathbf{A} \rangle &= \langle \psi | \mathbf{A} \otimes \mathbf{I}_B | \psi \rangle = \sum_{\mu, \nu} c_{\mu\nu}^* (\langle \mu |_A \otimes \langle \nu |_B) (\mathbf{A} \otimes \mathbf{I}_B) \sum_{i,j} c_{ij} (|i\rangle_A \otimes |j\rangle_B) \\ &= \sum_{i, \mu, j} c_{\mu j}^* c_{ij} \langle \mu | \mathbf{A} | i \rangle = \text{tr}(\mathbf{A} \rho_A) \end{aligned} \quad (1.13)$$

recovering (1.10). Here we used orthogonality for kets of the basis of system  $B$ .

Note that from the last line we can build up an expression for the density operator of the subsystem  $A$

$$\rho_A = \sum_{i, \mu, j} c_{\mu j}^* c_{ij} |i\rangle \langle \mu| \equiv \text{tr}_B (|\psi\rangle \langle \psi|) \quad (1.14)$$

Here, we find  $\rho_A$  by performing a *partial trace* over subsystem  $B$  of the density operator of the whole  $AB$  (in this case, a ray). That is, a linear map that takes an operator  $\mathbf{M}_{AB}$  on  $\mathcal{H}_A \otimes \mathcal{H}_B$  to an operator on  $\mathcal{H}_A$ .

### Purification

We have already seen that a mixed state of any quantum system can be realized as an ensemble of pure states (1.6). Moreover, this could happens in an infinite number of different ways, all of which have exactly the same consequences for any conceivable observation of the system. Therefore, although the preparation of a pure state is unambiguous (for the state  $\rho = |\psi\rangle \langle \psi|$ , measurement of the projection  $\mathbf{E} = |\psi\rangle \langle \psi|$  is guaranteed to have outcome one), for mixed state it is always ambiguous ([5]: section 2.5). Later, we will study how ambiguous it can be.

In addition, a mixed quantum state on a given quantum system described by a Hilbert space  $\mathcal{H}_A$ , can be always represented as the partial trace of a pure quantum

state on a larger bipartite system  $\mathcal{H}_A \otimes \mathcal{H}_B$ . We call this pure quantum state the *purification* of  $\rho_A$ . In fact, by considering a generic (possibly mixed) quantum state of the form (1.6) denoted by  $\rho_A$  (where the  $|a_i\rangle$  are not necessarily mutually orthogonal, this time), a bipartite pure state  $|\phi\rangle_{AB}$  which purifies  $\rho_A$  is of the form

$$|\phi\rangle_{AB} = \sum_i \sqrt{w_i} |a_i\rangle_A \otimes |b_i\rangle_B \quad (1.15)$$

where vectors  $|b_i\rangle_B \in \mathcal{H}_B$  are mutually orthonormal. Then, from (1.14) with (1.15), follows that

$$\rho_A = \text{tr}_B(|\phi\rangle_{AB} \langle\phi|_{AB}) \quad (1.16)$$

There are infinitely many purifications of a given mixed state, as the space  $\mathcal{H}_B$  and the basis  $\{|a_i\rangle\}$  and consequently  $\{|b_i\rangle\}$  can be chosen arbitrarily. All of this purifications only differ by a unitary transformation acting on  $\mathcal{H}_B$

$$|\phi'\rangle_{AB} = (\mathbf{I} \otimes \mathbf{U}_B) |\phi\rangle_{AB} \quad (1.17)$$

where we recall (1.12). Given a purification  $|\phi\rangle_{AB}$  of  $\rho_A$ , a measurement in system  $B$  that projects onto the  $\{|b_i\rangle\}$  basis, produces outcome  $|b_i\rangle$  with probability  $w_i$ , and will prepare with the same probability the pure state  $|a_i\rangle \langle a_i|$  of system  $A$ , realizing one ensemble interpretation of  $\rho_A$ . Moreover, from the relation (1.17), we can choose a purifying system  $|\phi\rangle_{AB}$ , such that anyone of the ensembles of the mixed state  $\rho_A$  can be realized by making different measurements on  $|\phi\rangle_{AB}$  (i.e., measuring a suitable observable of  $B$ , as in ). That is the *HJW* theorem ([5]: section 2.5.5).

### 1.2.3 Qubits

The fundamental unit of classical information is the *bit*, which takes only two possible values  $\{0,1\}$ . The corresponding unit of quantum information is the 'quantum bit' or *qubit*, which is a vector in a two-dimensional  $\mathcal{H}$ . According to the convention, we represent the elements of an orthonormal basis in this space as  $|0\rangle$  and  $|1\rangle$ . Therefore, the general state of a qubit, according to quantum mechanics, is usually a coherent superposition of both

$$|\psi\rangle = a |0\rangle + b |1\rangle, \quad |a|^2 + |b|^2 = 1, \quad a, b \in \mathbb{C} \quad (1.18)$$

According to (1.3), a measurement will projects the state onto the kets of the basis, and the outcome is not deterministic. For example, the probability that we obtain exactly the result  $|0\rangle$  is  $|a|^2$ . Therefore, the measurement irrevocably disturbs the state by destroying its coherence, except in the cases  $a = 0$  and  $b = 0$ . That means there is no way we can recover both the values of  $a$  and  $b$  after we performed a measurement.

In this regard, a qubit differs from a classical bit. In fact, we can measure a classical bit without disturbing it and deciphering all information it encodes by

measuring it only once. Indeed, by exploiting the *geometrical* interpretation of a qubit state (1.18), it is natural to interpret it as the spin state of an object with spin- $\frac{1}{2}$  ([5], subsection 2.2.1). Then  $|0\rangle$  and  $|1\rangle$  become the spin up  $|\uparrow\rangle$  and spin down  $|\downarrow\rangle$  states along a particular axis. Moreover, its quantum state is characterized by a unit vector  $\hat{n}$ , the spin's direction in a three-dimensional space (we say, a *Cartesian* space). Therefore, from the coefficients  $a$  and  $b$  results the orientation of the spin in this three-dimensional space (as they encode the polar angle  $\theta$  and the azimuthal angle  $\phi$ ). We recover that coefficients' relative phase also has a physical significance and that we can determine  $|\psi\rangle$  only by measuring along *both* the spin axis. In fact, that is equivalent to determine the unit vector  $\hat{n}$ : altogether measurements along  $x$ ,  $y$ ,  $z$  are required.

### Bloch sphere and maximally mixed state

Exploring deeply the geometrical representation, in (1.18) we have the complex coefficients  $a$  and  $b$ . Therefore, we expect  $|\psi\rangle$  to have four degree of freedom. However, the relation between  $a$  and  $b$  in (1.18) gets one. This means, there must exist a suitable change of coordinates so that the system of coefficients has only three degree of freedom. The following expression fulfils the requirement (these are the *Hopf coordinates*) ([5]: section 2.3.2).

$$a = e^{i\delta} \cos \theta/2 \qquad b = e^{i(\delta+\phi)} \sin \theta/2 \qquad (1.19)$$

Because  $e^{i\delta}$  has no physical observable consequences (remind (1.3) and  $|e^{i\delta}|^2 = 1$ ), and by arbitrary choose for  $a$  to be real, we obtain

$$|\psi\rangle = \begin{pmatrix} \cos \theta/2 \\ e^{i\phi} \sin \theta/2 \end{pmatrix} = \begin{pmatrix} e^{-i\phi/2} \cos \theta/2 \\ e^{i\phi/2} \sin \theta/2 \end{pmatrix} \qquad (1.20)$$

From this we may explicitly compute

$$\boldsymbol{\rho} = |\psi\rangle \langle \psi| = \begin{pmatrix} \cos^2 \theta/2 & \cos \theta/2 \sin \theta/2 e^{-i\phi} \\ \cos \theta/2 \sin \theta/2 e^{i\phi} & \sin^2 \theta/2 \end{pmatrix} = \frac{1}{2} (\mathbf{I} + \hat{n} \cdot \vec{\boldsymbol{\sigma}}) \qquad (1.21)$$

were  $\hat{n} = (\sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta)$  and  $\vec{\boldsymbol{\sigma}}$  is a vector whose components are the traceless Pauli matrices. In fact, every single qubit system can be described by a  $2 \times 2$  self-adjoint matrix that can be expanded in the basis  $\{\mathbf{I}, \vec{\boldsymbol{\sigma}}_1, \vec{\boldsymbol{\sigma}}_2, \vec{\boldsymbol{\sigma}}_3\}$  ([5]: section 2.3.2). That is called the *Bloch* parametrization for qubit pure states, that can be represented by any point on the unit surface. Notice that, as (1.6) holds, for mixed states we have that  $\boldsymbol{\rho}^2 = \sum_i w_i^2 < \sum_i w_i = 1$ . Therefore,  $\hat{n}$  can be replaced with a vector  $\vec{P}$  so that  $|\vec{P}|^2 < 1$ . Consequently, also mixed states can be represented with internal points on the unit three-dimensional ball called *Bloch Sphere*. We see a perfect correspondence between the density matrix of a qubit and the points contained into the *Bloch sphere*, with  $0 \leq |\vec{P}| < 1$  for mixed states (internal points) and  $|\vec{P}| = 1$  for pure states (surface points).

It is now become easy to answer the question about how ambiguous the preparation of mixed states could be. In fact, the more a state is mixed, the more it is ambiguous its preparation. In order to comprehend this, notice that from the properties of  $\rho$  it follows immediately that, given two density matrices  $\rho_1$  and  $\rho_2$ , we can always construct another density matrix as a convex linear combination of the two

$$\rho(\lambda) = \lambda\rho_1 + (1 - \lambda)\rho_2 \quad (1.22)$$

For a point  $0 < |\vec{P}| < 1$  interior to the Bloch sphere, we have

$$\rho(\vec{P}) = \lambda\rho(\hat{n}_1) + (1 - \lambda)\rho(\hat{n}_2) \quad (1.23)$$

if  $\vec{P} = \lambda\hat{n}_1 + (1 - \lambda)\hat{n}_2$ . In other words, if  $\vec{P}$  lies somewhere on the line cord with extremal points the pure states indicated by  $\hat{n}_1$  and  $\hat{n}_2$  ([5]: section 2.5.1). It follows that, the more  $\vec{P}$  is 'internal' ( $\hat{n}_1$  and  $\hat{n}_2$  are close to be opposed), the more are the different ways trough  $\rho(\vec{P})$  can be expressed as pure states combination. Consequently, from what said and (1.21), we have that the maximal ambiguity preparation is reached for the limit situation of the *maximally mixed* state of a single qubit

$$\rho = \frac{1}{2}\mathbf{I} \quad (1.24)$$

In this case  $|\vec{P}| = 0$  and this state can be prepared as an ensemble of pure states in an infinite variety of ways. Therefore, in (1.24),  $\rho$  conveys no information at all about the state preparation. The *purity* of a state can be visualized as the degree in which it is close to the surface of the sphere. It is denoted with  $\text{tr}(\rho^2) \leq 1$  and it is one only for pure states.

### 1.2.4 Schmidt decomposition and Maximal Entanglement

As we could notice, one of the situations which cause mixed states to arise is entanglement between two systems. We already discussed about entanglement, but now we approach this argument from another point of view. It is, in fact, possible to decompose a bipartite pure state, like (1.4), in a very usefull form.

We can rewrite a vector in  $\mathcal{H}_A \otimes \mathcal{H}_B$  using its *Schmidt decomposition*

$$|\psi\rangle_{AB} = \sum_i \sqrt{w_i} |a_i\rangle_A \otimes |b'_i\rangle_B \quad (1.25)$$

were  $w_i$  are the same probabilities as in (1.6). The orthonormal schmidt basis  $|a_i\rangle_A$  is chosen to be the basis in which  $\rho_A$  is diagonal. The other orthonormal basis  $|b'_i\rangle_B$  is obtained by comparing (1.6) with (1.7), resulting from the partial trace over  $B$  performed with respect to (1.4) ([5]: section 2.4)

$$|b'_i\rangle_B = w_i^{-1/2} \sum_j c_{ij} |b_j\rangle_B$$

where we always consider at least one  $i$  to perform Schmidt decomposition, so that  $w_i \neq 0$ . It is possible to say if a state is entangled by studying its Schmidt decomposition. The strictly positive values  $\sqrt{w_i}$ , in (1.25), are the *Schmidt coefficients*. The number of Schmidt coefficients, counted with its multiplicity, is the *Schmidt number*. Taking up the statement that if a sum like (1.6) has more than one terms, then it would describe a mixed state, we can see that a pure state  $|\psi\rangle_{AB}$  is entangled if and only if its Schmidt number is greater than one. Otherwise, it is separable and its subsystems  $A$  and  $B$  are pure. It turns out that, by increasing the Schmidt number of a state, we create entanglement. However, the Schmidt number is preserved under local unitary transformations on system  $A$  or system  $B$  alone. therefore, *entanglement cannot be created locally* ([5]: section 2.4).

Another important result coming from Schmidt decomposition is that, by considering  $\rho = |\psi\rangle\langle\psi|_{AB}$ , its partial trace, with respect to either system  $A$  or  $B$ , is a diagonal density matrix whose nonzero diagonal elements are  $w_i$ . In other words,  $\rho_A$  and  $\rho_B$  have the same nonzero eigenvalues and a different number of zero eigenvalues only if  $\dim(\mathcal{H}_A) \neq \dim(\mathcal{H}_B)$ . It is possible to recover the Schmidt decomposition of  $|\psi\rangle_{AB}$  by diagonalizing  $\rho_A$  and  $\rho_B$  and then pairing up the eigenstates  $|a_i\rangle_A$  and  $|b'_i\rangle_B$  which share the same eigenvalue  $w_i$ . If  $\rho_A$  has no degenerate nonzero eigenvalues, then there is a unique decomposition for  $|\psi\rangle_{AB}$ . Conversely, if degeneration for nonzero eigenvalues of  $\rho_A$  occurs, we need more information about which  $|b'_i\rangle_B$  gets paired with each  $|a_i\rangle_A$ , causing ambiguity. In fact, there will exist unitary change of Schmidt basis that will give a valid decomposition. In the maximal degeneration case, in which  $\rho_A$  and  $\rho_B$  are *maximally mixed*, so that schmidt coefficients are all equal to one, the state  $|\psi\rangle_{AB}$  is said to be *maximally entangled*. Therefore, unitary transformation will give a valid decomposition ([5]: section 2.4). In general, the Schmidt decomposition  $|\psi\rangle_{AB}$  corresponds to the *purification* (1.15) of  $\rho_A$  in the special case in which the  $|a_i\rangle$  are orthonormal. So, in the case of maximal degeneration, we recover the *HJW theorem* as a simple corollary of Schmidt decomposition ([5]: section 2.5.5). Considering  $\dim(\mathcal{H}_A) = \dim(\mathcal{H}_B) = d$ , if we have  $d$  terms in the sum (1.25) all the  $w_i$  will be equal to  $\frac{1}{\sqrt{d}}$ , then, fixed a schmidt basis, the state  $|\psi\rangle_{AB}$  is said to be *maximally entangled* and it will yield maximally mixed operators  $\rho_A = \rho_B = \frac{1}{d}\mathbf{I}$  when take partial traces.

### 1.2.5 Fidelity of Quantum States

It is possible to measure the distinguishability of two quantum states as the deviation from one of their *fidelity*. For two pure states,  $|\phi\rangle$  and  $|\psi\rangle$ , the fidelity is given by their overlap

$$F(|\phi\rangle, |\psi\rangle) = |\langle\phi|\psi\rangle|^2 \quad (1.26)$$

In fact, it can be thought as the probability that an input state  $|\phi\rangle$  passes the ' $|\psi\rangle$ ' test, which is the measurement in the basis  $|\psi\rangle, |\psi^\perp\rangle$ . More in general, for two



density operator,  $\rho$  and  $\sigma$ , the fidelity is defined by ([5]: section 2.6.1)

$$F(\rho, \sigma) \equiv \left( \text{tr} \sqrt{\rho^{\frac{1}{2}} \sigma \rho^{\frac{1}{2}}} \right)^2 = (\text{tr} \sqrt{\rho \sigma})^2 \quad (1.27)$$

where the last equality results from properties of the unique positive square root for a positive semidefinite matrix. From (1.27) and using the Bloch parametrization, we can obtain an explicit formula for the fidelity in the case of two states of a qubit with polarizations  $\vec{P}$  and  $\vec{Q}$  (1.21) ([6]: section 3.8, result from exercise 3.7)

$$F(\vec{P}, \vec{Q}) = \frac{1}{2} \left( 1 + \vec{P} \cdot \vec{Q} + \sqrt{(1 - \vec{P}^2)(1 - \vec{Q}^2)} \right) \quad (1.28)$$

Considering pure qubit states  $|\phi\rangle$  and  $|\psi\rangle$ , we assign them polarizations  $\vec{P}$  and  $\vec{Q}$ . Until now, the state  $|\phi\rangle$  is somewhere on the Bloch sphere. We might as well orient the sphere so that his direction  $\vec{P}$  became the  $\hat{z}$  direction. With an angle  $\theta$  between the two states on the Bloch sphere, from (1.28) we have  $F(\vec{P}, \vec{Q}) = \frac{1}{2}(1 + \cos \theta)$ . Moreover, on the *average*, a guess state  $|\phi\rangle$  will match  $|\psi\rangle$  with fidelity

$$\langle F \rangle = \frac{1}{4\pi} \int_0^{2\pi} d\phi \int_0^\pi d\theta \sin \theta \frac{1}{2} (1 + \cos \theta) = \frac{1}{2} \quad (1.29)$$

In order to improve the guess, we might make a measurement of  $|\phi\rangle$ , say, along the  $\hat{z}$  axis. Given the result  $k \in \{0, 1\}$ , our guess is then the state  $|k\rangle$ . The fidelity thus becomes  $F_k(|k\rangle, |\psi\rangle) = |\langle k|\psi\rangle|^2$ , depending on the value of  $k$ , causing this to occur with probability  $p_k = |\langle k|\psi\rangle|^2$ . So, averaging over all possible outcomes for  $k$ , we have  $F = \sum_k p_k F_k = \sum_k |\langle k|\psi\rangle|^4$ . Therefore, averaging over a generic  $|\psi\rangle = \cos \frac{\theta}{2} |0\rangle + e^{i\phi \sin \frac{\theta}{2}} |1\rangle$  from (1.20), the fidelity is

$$\langle F \rangle = \frac{1}{4\pi} \int_0^{2\pi} d\phi \int_0^\pi d\theta \sin \theta \left( \cos^4 \frac{\theta}{2} + \sin^4 \frac{\theta}{2} \right) = \frac{2}{3} \quad (1.30)$$

Thus making the measurement increases the fidelity of the guess.

### 1.2.6 Generalized measurements and POVM

We already introduced PVM performed on a subsystem  $A$  of larger system. However, we can perform a PVM on the whole  $AB$ . In that case, the effect on  $A$  alone need not be an orthogonal projection.

We start considering  $B$  to be an *ancilla* system of  $A$  (this name means  $B$  is used to achieve a desired goal). We also define set of orthogonal projectors  $\{\mathbf{E}_i, i \in O\}$  as in subsection 1.2.2. Performing a unitary transformation  $\mathbf{U}$  on  $AB$ , such as a time evolution operation, we entangle the two subsystems. So, by observing the ancilla in its fiducial basis, we can perform any conceivable orthogonal measurement on the system  $A$ . However, in general the unitary transformation could picks out a different



preferred basis than the fiducial one. Then the measurement of  $B$  causes non orthogonal states of  $A$ . Such a measurement, is said *generalized* and it is powered by some special nonnegative operators, called *POVM* (Positive Operator-Valued Measure). That is, a complete set of positive semi-definite Hermitian matrices  $\{\mathbf{F}_i, i \in O\}$  on a Hilbert space  $\mathcal{H}_A$ . We also introduce the *Kraus* operators  $\{\mathbf{M}_i, i \in O\}$  so that

$$\mathbf{F}_i = \mathbf{M}_i^\dagger \mathbf{M}_i \quad (1.31)$$

Being  $|\phi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$  and  $|\psi\rangle \in \mathcal{H}_A$ , by expanding the action of  $\mathbf{U}$  in the basis for  $B$ , we have

$$\mathbf{U} : |\phi\rangle_{AB} = |\psi\rangle_A \otimes |0\rangle_B \mapsto |\phi'\rangle_{AB} = \sum_i \mathbf{M}_i |\psi\rangle_A \otimes |i\rangle_B \quad (1.32)$$

Since  $\mathbf{U}$  is unitary, we have  $\sum_i \mathbf{M}_i^\dagger \mathbf{M}_i = 1$ . Remember that  $\mathbf{M}_a \mathbf{M}_b = \delta_{ab} \mathbf{M}_a$  does not hold in general, so, in contrast to an orthogonal measurement, the post-measurement states are typically not orthogonal. If we are considering that subsystem  $A$  is described by  $\rho_A$ , in the *Heisenberg picture* we have that a POVM arises when a PVM is performed on  $AB$ . In fact the relation

$$\mathbf{F}_i = \mathbf{U}^\dagger \mathbf{E}_i \mathbf{U} \quad (1.33)$$

holds [12]. The Kraus decomposition (1.31) is unique unless unitary operator  $\mathbf{U}_i$ . In this case also holds that

$$\mathbf{M}_i = \mathbf{U}_i \sqrt{\mathbf{F}_i} \quad (1.34)$$

Therefore, the probability of obtaining outcome  $i$  with this PVM, and the state suitably transformed by the unitary, is the same as the probability of obtaining it with the original POVM

$$\begin{aligned} \text{Prob}(i) &= \text{tr}(\mathbf{U}^\dagger \rho_A \mathbf{U} \mathbf{E}_i) = \text{tr}(\rho_A \mathbf{U}^\dagger [\mathbf{I}_A \otimes |i\rangle \langle i|_B] \mathbf{U}) \\ &= \text{tr}\left(\rho_A \left(\sum_i \sqrt{\mathbf{F}_{j_A}}^\dagger \otimes \langle j|_B\right) [\mathbf{I}_A \otimes |i\rangle \langle i|_B] \left(\sum_i \sqrt{\mathbf{F}_{k_A}} \otimes |k\rangle_B\right)\right) \\ &= \text{tr}\left(\rho_A \sqrt{\mathbf{F}_{i_A}} \mathbf{I}_A \sqrt{\mathbf{F}_{i_A}}\right) = \text{tr}(\rho_A \mathbf{F}_i) \end{aligned} \quad (1.35)$$

where we obtained the expression of  $\mathbf{U}$  from (1.32), the expression for  $\mathbf{E}_i$  from (1.11) and we used (1.34). Furthermore, if two such measurements are performed in rapid succession, the outcomes need not to be the same. Moreover, the post-measurement state corresponding to outcome  $i$  is arbitrary, since we are free to choose the unitary  $\mathbf{U}_i$  in (1.34) however we please for each possible outcome ([6]: section 3.1.2).

### No cloning theorem

Another interesting thing regarding the non orthogonal states, is that they cannot be distinguished without being disturbed. More explicitly, there not exists, in general, a

quantum Xerox machine; that is the *No cloning theorem* [2], widely used in Quantum Cryptography.

It deals with the fact that does not exist any unitary  $\mathbf{U}$  in  $\mathcal{H}_A \otimes \mathcal{H}_B$ , such that

$$\mathbf{U}(|\phi\rangle_A |e\rangle_B) = e^{i\theta(\phi,e)} |\phi\rangle_A |\phi\rangle_B \quad (1.36)$$

where  $|\phi\rangle$  and  $|e\rangle$  are all normalized states and  $\theta$  is a real number depending on the two states. In fact, due to unitarity of  $\mathbf{U}$ , it holds that

$$\langle\psi|\phi\rangle \langle e|e\rangle = e^{i[\theta(\phi,e)-\theta(\psi,e)]} \langle\psi|\phi\rangle^2 \quad (1.37)$$

Since  $|e\rangle$  is assumed to be normalized, we have

$$|\langle\psi|\phi\rangle| = |\langle\psi|\phi\rangle|^2 \quad (1.38)$$

This implies that either  $|\langle\psi|\phi\rangle| = 1$  or  $|\langle\psi|\phi\rangle| = 0$ , causing the existence of a unitary like  $\mathbf{U}$  only for states which represent the same ray or orthogonal states. Hence, no unitary machine can make a copy of both  $\psi$  and  $\phi$  if they are *distinct non orthogonal* states. Therefore, a single universal  $\mathbf{U}$  cannot clone a *general* quantum state. This proves the no cloning theorem.

### 1.2.7 Quantum channels

If a state of a bipartite system undergoes unitary evolution, we describe the evolution of  $A$  alone by a linear map  $\varepsilon$  called *quantum channel*. We may imagine to measure  $B$  in its basis, but failing to record the outcome, so we are forced to average over all of the possible post-measurement states, weighted by their probabilities ([6]: section 3.2.1). Therefore, the result for a state  $\rho_A$  of  $A \in \mathcal{H}_A$  is a linear map, so that

$$\varepsilon(\rho_A) = \sum_i M_i^\dagger \rho_A M_i \quad (1.39)$$

It easy to verify that  $\varepsilon$  is a linear map that preserves hermiticity, positivity and trace. Quantum channels are important in giving formalism to discuss *decoherence* [6], the evolution of pure states into mixed ones. In fact, if the sum (1.39) has only one term, then the evolution of  $\rho_A$  is unitary and the channel is said *pure*. Otherwise, the channel transforms pure initial states of  $A$  to mixed ones ([6]: section 3.2.1). For example, in (1.35) we see that the operation  $\mathbf{U}^\dagger \rho_A \mathbf{U}$  induces a non-pure channel, as the time evolution operator is unitary only if acting on the whole  $AB$ . In particular, exploiting the action of the channel, we have that pure states of  $A$  become entangled with  $B$  under the joint unitary transformation described by (1.32), and therefore the state of  $A$  becomes mixed when we trace out  $B$ . In (1.33), we notice the same thing, but represented in the Heisenberg picture, where states are stationary and operators evolve instead. The maps induced is the *dual*  $\varepsilon^*$  of  $\varepsilon$  ([6]: section 3.2.3). For a general operator  $\mathbf{A}$ , we have

$$\varepsilon^*(\mathbf{A}) = \sum_i M_i^\dagger \mathbf{A} M_i \quad (1.40)$$

Another name for  $\varepsilon$  is TPCP map, which means trace-preserving completely positive map. If an ancilla  $B$  of arbitrary finite dimension  $n$  is coupled to the system  $A$ , then the induced map  $\varepsilon \otimes \mathbf{I}_n$ , where  $\mathbf{I}_n$  is the identity map on the ancilla, must also be positive. Therefore, for complete positivity to hold it is required that  $\varepsilon \otimes \mathbf{I}_n$  is positive for all  $n$ . We say  $\varepsilon \otimes \mathbf{I}_n$  maps  $AB$  to  $A'B$ .

### Channel-state duality and relative state method

The evolution of  $A$  is in general non-unitary. We are therefore entitled to imagine that  $A$  is a part of an extended system which evolves unitarily. Such an evolution law can be encoded in a quantum channel. In fact every channel has its unitary representation ([6]: section 3.2). To study this, we introduce the ancilla  $B$  having the same dimension  $d$  as  $A$ . Therefore  $\varepsilon$ , being completely positive, maps a maximally entangled state on  $AB$  to a density operator on  $A'B$ . So, we have

$$|\tilde{\phi}\rangle_{AB} = \sum_{i=0}^{d-1} |i\rangle_A \otimes |i\rangle_B \quad (1.41)$$

where we used unconventional normalization for the sake of simplicity (the tilde denotes that). We also have

$$(\varepsilon \otimes \mathbf{I}_n) |\tilde{\phi}\rangle \langle \tilde{\phi}|_{AB} = \sum_a |\tilde{\psi}_a\rangle \langle \tilde{\psi}_a|_{A'B} \quad (1.42)$$

where the probability of each state are absorbed by the normalization. Conversely, we may associate with any density operator on  $A'B$  a corresponding channel taking  $A$  to  $A'$ . This is the *Choi-Jamiolkowski isomorphism* or *channel-state duality*. To verify this, we notice that

$$|\varphi\rangle_A = \sum_i c_i |i\rangle_A = \sum_i c_i ({}_B \langle i | \tilde{\phi} \rangle_{AB}) = {}_B \langle \varphi^* | \tilde{\phi} \rangle_{AB} \quad (1.43)$$

Then the channel is recovered from the density matrix of  $A'B$

$$\varepsilon ({}_A \langle \varphi | \varphi \rangle_A) = \sum_a ({}_A \langle \varphi^* | \tilde{\psi}_a \rangle \langle \tilde{\psi}_a | \varphi^* \rangle)_{A'} \quad (1.44)$$

after defining the Kraus operators as

$$\mathbf{M}_a |\varphi\rangle_A = {}_B \langle \varphi^* | \tilde{\psi}_a \rangle_{A'B} \quad (1.45)$$

This scheme for extracting the action on  $|\varphi\rangle_A$  using its dual  $\langle \varphi^* |_B$  is called the *relative state method*.

### 1.2.8 Axioms revisited

With the theory of QoS, we can rightly rewrite axioms 1, 3 and 4 of (1.1) in a more general formulation:

- (1) A *state* is a density operator in  $\mathcal{H}$ ;
- (3) A *measurement* is a *positive operator-valued measure (POVM)*;
- (4) *Time evolution* is described by a *trace-preserving completely positive map (TPCP map)*.

## 1.3 Entanglement-assisted communication

### 1.3.1 Hidden Quantum Information and *LOCC*

Quantum entanglement features quantum information and it can be established under specific condition.

We start considering what a maximally entangled state of two qubits is

$$|\phi^+\rangle_{AB} = \frac{1}{\sqrt{2}} (|00\rangle_{AB} + |11\rangle_{AB}) \quad (1.46)$$

Tracing over qubit  $B$  we obtain a multiple of the identity operator that has every state as an eigenstate. Therefore, any local measurement on  $A$  or  $B$  will generate a random bit giving no information about the preparation. Conversely, in the case of a single qubit we could store a bit by preparing, say, either  $|\uparrow_{\hat{n}}\rangle$  or  $|\downarrow_{\hat{n}}\rangle$  and recover it by measuring the qubit along the  $\hat{n}$ -axis. We would, therefore, be able to encode two bits of classical information in two qubits, but in  $|\phi^+\rangle$  this information seems to remain *hidden* by measuring  $A$  or  $B$ . In fact, consider  $|\phi^+\rangle$  to be one member of a basis of four mutually orthogonal and maximally entangled states for two qubits

$$|\phi^\pm\rangle = \frac{1}{\sqrt{2}} (|00\rangle \pm |11\rangle) \quad |\psi^\pm\rangle = \frac{1}{\sqrt{2}} (|01\rangle \pm |10\rangle) \quad (1.47)$$

If we encode one bit as the *parity* bit ( $|\phi\rangle$  or  $|\psi\rangle$ ) and the other as the *phase* bit (+ or -), then all we can do locally is to manipulate this information by applying  $\sigma_1$  or  $\sigma_3$  in one member of the entangled pair. Indeed, this will flip the phase bit or the parity bit stored in the whole entangled state, respectively ([7]: section 4.1.1). In fact, as for *HJW theorem*, we can always perform *local* unitary transformation that changes one maximally entangled state to any other maximally entangled state (only acting on  $\mathcal{H}_A$  or  $\mathcal{H}_B$  separately, NOT on  $\mathcal{H}_A \otimes \mathcal{H}_B$ ). Nevertheless, remains that neither one can read this information. In fact, it is possible to infer one bit encoded in the entangled state by measuring the same bit on each qubit separately. But, it is impossible to recover one bit in the entangled state without disturbing

the other one. Indeed, the entangled basis states are common eigenstates of two commuting observables whose eigenvalues are the phase bit and the parity bit

$$\sigma_1^{(A)} \otimes \sigma_1^{(B)} \qquad \sigma_3^{(A)} \otimes \sigma_3^{(B)} \qquad (1.48)$$

These operators commute, but they cannot be measured simultaneously by performing local measurement on both qubits separately. In fact, we have both  $\sigma_i^{(A)}$  and  $\sigma_i^{(B)}$  commute with  $\sigma_i^{(A)} \otimes \sigma_i^{(B)}$ , but not with  $\sigma_j^{(A)} \otimes \sigma_j^{(B)}$ , where  $i \neq j$  and  $i, j = 1, 3$ . ([7]: section 4.1.1). Despite this, there exists a way to rotate the entangled basis (1.47) to unentangled basis  $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$  and then measuring both bits we can learn the value of both the phase bit and parity bit encoded in  $|\phi^+\rangle_{AB}$ . The upshot is that exists a transformation that establishes or removes entanglement (by running it backward). This transformation is composed by a *local* part, that is a unitary  $\mathbf{H}$  performed on the first qubit (*Hadamard transform*)

$$\mathbf{H} = \frac{1}{\sqrt{2}} (\sigma_1 + \sigma_3) \qquad (1.49)$$

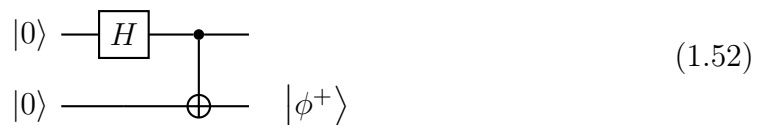
and by a *nonlocal* part  $\mathbf{CNOT}$  that removes or establishes entanglement and requires both qubits are in the same position at the same time to act (the *controlled-NOT transformation*)

$$\mathbf{CNOT} : |a, b\rangle \rightarrow |a, a \oplus b\rangle \qquad (1.50)$$

where  $a, b = 0, 1$  are the basis states and  $a \oplus b$  denotes addition modulo 2. In fact, we see that for a state  $|00\rangle$  the transformation acts in two steps, applying  $\mathbf{H}$  to the first qubit and eventually flips the second qubit depending on the outcome of  $\mathbf{H}$  (this performed by  $\mathbf{CNOT}$ )

$$|00\rangle \rightarrow \frac{1}{\sqrt{2}} (|0\rangle + |1\rangle) |0\rangle \rightarrow |\phi^+\rangle \qquad (1.51)$$

Hence, an entangled state is obtained. By using the quantum circuit notation (to be read from left to right), where qubits are denoted as horizontal lines, we have



$$\begin{array}{c} |0\rangle \text{---} \boxed{H} \text{---} \bullet \text{---} \\ |0\rangle \text{---} \oplus \text{---} \end{array} \quad |\phi^+\rangle \qquad (1.52)$$

Similarly, we can reverse the circuit and transform an entangled two-qubit state into an unentangled one.

From all of this examples, we see that *local operations and classical communications (LOCC)* will not suffice for create entanglement among distantly separated pairs, or extract information encoded in it.

### 1.3.2 Dense Coding and Quantum Teleportation

Although by using *LOCC* is impossible to reveal all the information encoded in entangled states, that could be used to enhance the transmission of classical information. This procedure is named *Dense Coding*. We start by introducing the two friends Alice and Bob who share a *quantum* channel. Alice want to send a message to Bob using qubits in place of classical bits (*cbits*), by preparing them, say, either  $|\uparrow_z\rangle$  or  $|\downarrow_z\rangle$  and then letting Bob to measure along  $\hat{z}$  to infer the choice she made. In first place, sending one qubit at a time, no matter how she prepares it and no matter how Bob measures it, no more than one cbit can be carried by each qubit (even in the qubits are entangled with one another) ([7]: section 4.4.1). Despite this, in some special conditions Alice can transmits two classical bits by sending a single qubit to Bob, exploiting Dense Coding. More specifically, it is necessary for them to share an entangled pair of qubits in a maximally entangled state, for example  $|\phi^+\rangle_{AB}$ . The maximal entanglement condition is needed to grant quantum channel to have perfect fidelity  $F = 1$  ([7]: section 4.4.3). However, this state must be prepared by either of them and later one one member of the entangle pair must be shipped to the other. Then, Alice and Bob actually need to use the channel twice to exchange two bits of information. Hence, also in Dense Coding we recover that each qubit convey a cbit of information. Anyway, the upshot is that the first qubit is shipped long before anyone knew what the message was going, making possible to send two cbits by only using the channel once when needed.

More in detail, Alice can transmits the second cbit of information by carrying out a specific *protocol* on her entangled qubit (*ebit*). She can perform one of four possible unitary transformations, consequently changing  $|\phi^+\rangle_{AB}$  to one of four mutually orthogonal states of (1.47). Respectively, we have

$\mathbf{I}$  (she does nothing)  $\rightarrow$  obtains  $|\phi^+\rangle_{AB}$

$\sigma_1$  (180° rotation about  $\hat{x}$ -axis)  $\rightarrow$  obtains  $|\psi^+\rangle_{AB}$

$\sigma_2$  (180° rotation about  $\hat{y}$ -axis)  $\rightarrow$  obtains  $|\psi^-\rangle_{AB}$  (up to a phase)

$\sigma_3$  (180° rotation about  $\hat{z}$ -axis)  $\rightarrow$  obtains  $|\phi^-\rangle_{AB}$

Now, she send her qubit to Bob, who receives and perform an orthogonal collective measurement on the pair that projects onto the maximally entangled basis. The measurement outcome unambiguously distinguishes which one of the four possible actions Alice committed. Therefore, a single qubit transmitted by Alice successfully carried two bits of classical information (the other cbits is encoded in the preparation of the qubit). An important feature of Dense Coding is that the shipper need not worry that an eavesdropper will intercept the transmitted qubit and decipher the message. In fact, the shipped qubit is a part of a maximally entangled state, therefore its density matrix is  $\rho_A = \frac{1}{2}\mathbf{I}_A$ , and so carries no information at all. All the information is encoded in entanglement between  $A$  and  $B$ , and therefore cannot

be recovered locally.

Surprisingly, it is also possible to perform the converse of Dense Coding, that is, send two classical bits to convey a qubit. This procedure is named *Quantum teleportation*. In fact, to measure a qubit along  $\vec{n}$ -axis and later tell the outcome to Bob so that he prepares his qubit accordingly, does not suffice for Alice to transmit her qubit. Indeed, by measuring along  $\vec{n}$ -axis, Bob's state will match Alice's qubit with a Fidelity given by (1.30). However, they can achieve Fidelity  $F = 1$  by following this protocol: Alice combines the qubit  $|\psi\rangle_C$  she wants to send to Bob with her half of the  $|\phi_{AB}\rangle$  pair. That is possible by measuring the *Bell states* of  $A$  and  $C$ , which means to measure the two commuting observables

$$\sigma_1^C \otimes \sigma_1^A, \quad \sigma_3^C \otimes \sigma_3^A \quad (1.53)$$

that project into one of the four maximally entangled states  $|\phi^\pm\rangle_{CA}$ ,  $|\psi^\pm\rangle_{CA}$ . In fact, in [7]: section 4.4.2, is calculated that

$$|\psi\rangle_C |\phi^+\rangle_{AB} = \frac{1}{2} |\phi^+\rangle_{CA} |\psi\rangle_B + \frac{1}{2} |\psi^+\rangle_{CA} \sigma_1 |\psi\rangle_B + \quad (1.54)$$

$$\frac{1}{2} |\psi^-\rangle_{CA} (-i\sigma_2) |\psi\rangle_B + \frac{1}{2} |\phi^-\rangle_{CA} \sigma_3 |\psi\rangle_B \quad (1.55)$$

Consequently, the results of Bell measurement are shipped to Bob, conveyed by two classical bits. This unequivocally select one of the four terms of (1.54). At this point, for (1.54), Bob knows the action he has to perform on a random qubit  $|\psi\rangle_B$  in order to transform it into a *perfect* copy of  $|\psi\rangle_C$ .

Finally, we note that the teleportation procedure is fully consistent with the no cloning principle, as for when a copy of  $|\psi\rangle_C$  appears on Bob's hands after he performed one of the four suitable unitaries, the original  $|\psi\rangle_C$  had to be destroyed by Alice's measurement which establishes correlation between  $A$  and  $C$ .





## Chapter 2

# Dawn of Quantum Information Theory

In this chapter we will be occupied with generalizing Claude Shannon's great classical contributions to a quantum settings. We start discussing about Shannon entropy and its relevance on classical information theory [1]. Moreover, we will see how it naturally deals with the problem of data compression in Shannon's information theory. So, we will study how much one can decrease the redundancy incorporated in the message (*source coding theorem*). In particular, entropy provides a suitable way to quantify redundancy.

One of the major thrusts of Shannon's Quantum Information Theory is to transmit classical and quantum information through noisy quantum channels. To study this, we start inspecting classical information theory and focusing on the *rate* that allows us to communicate reliably over a noisy channel (*noisy channel coding theorem*). We will always consider an asymptotic setting when considering features of information theory. So, the same quantum channel or state is used many times, ignoring practical issues.

### 2.1 Shannon Theory

When we refer to a "message", we mean a string of letters, where each of them is chosen from an alphabet of  $k$  possible letters. Consider that each letter  $x$  is picked up by sampling from a probability distribution  $\mathcal{X}$ , that associates a probability  $p(x)$  to each letter

$$\mathcal{X} = \{x, p(x)\} \tag{2.1}$$

where  $0 \leq p \leq 1$  holds. Since the letter are statistically independent and identically distributed, the particular string of letters or bits  $\vec{x} = \{x_1 x_2 \dots x_n\}$  occurs with probability

$$p(x_1 x_2 \dots x_n) = \prod_{i=1}^n p(x_i) \tag{2.2}$$

Hence, for  $n$  very large, every typical string will contain the letter  $x$  about  $np(x)$  times. A string is said *atypical* if  $p = \frac{1}{2}$ , while for every other allowed values of  $p$  it is said *typical*. The number of distinct strings is given by the multinomial coefficient  $\frac{n!}{\prod_x (np(x))!}$ , and, from the *Stirling approximation*  $\log_2 n! \approx n \log_2 n - n \log_2 e$ , we have

$$\log_2 \frac{n!}{\prod_x (np(x))!} \approx nH(\mathcal{X}) \quad \text{with} \quad H(\mathcal{X}) = - \sum_x p(x) \log_2 p(x) \quad (2.3)$$

where logarithms with base 2 are more convenient for expressing a quantity of information in bits. Slightly above,  $H(\mathcal{X})$  is the *Shannon entropy* of the ensemble with probability distribution given by (2.1). Therefore, the number of typical strings is of order  $2^{nH(\mathcal{X})}$ . To convey all the information carried by a string of  $n$  bits, it suffices to choose a block code that assigns a nonnegative integer to each of the typical  $n$  letter strings. So, it needs to distinguish about  $2^{nH(\mathcal{X})}$  messages and every information in a string of  $n$  letters can be expressed using a string of  $nH(\mathcal{X})$  bits. In this sense a letter  $x$  chosen from the ensemble carries, on the average,  $H(\mathcal{X})$  bits of information. Since  $0 \leq H(\mathcal{X}) \leq 1$  for  $0 \leq p(x) \leq 1$ , and  $H(\mathcal{X}) = 1$  only if  $p(x) = \frac{1}{2}$ , the block code shorten the message for any typical sequence. Information is therefore compressed. To conclude, the probability for a message to be atypical becomes negligible asymptotically, for very large values of  $n$  ([8]: section 10.1.1).

From what said, we are now able to generalize the concept of typical strings to  $\delta$ -typical strings. Still remain  $n$ -letter strings, with large values of  $n$ . These, are featured by a probability  $p(\vec{x})$  satisfying

$$2^{-n(H+\delta)} \leq p(\vec{x}) \leq 2^{-n(H-\delta)} \quad (2.4)$$

being  $\delta > 0$ . Therefore, all sequences can be encoded in a block code with length  $n(H + \delta)$ . For large  $n$ , thanks to the Asymptotic Equipartition Property (demonstrated in [13]), a  $\delta$ -typical string tend to become a typical one, so the probability for a given sequence to be  $\delta$ -typical become arbitrary close to one

$$p_{\text{success}} \geq 1 - \epsilon \quad (2.5)$$

where  $\epsilon > 0$  and is small. Conversely, we want to compress even further and we fix a positive constant  $\delta'$ , so that every message is encoded in  $H - \delta'$  bits. We see that only  $2^{n(H-\delta')}$  typical messages with probability no higher than  $2^{-n(H-\delta)}$  can be correctly decoded. However, we can make  $\epsilon$  and  $\delta$  small as we please, so  $p_{\text{success}}$  becomes very small for  $n \rightarrow \infty$ . That is Shannon's *source coding theorem* which asserts that a compression rate  $H(\mathcal{X}) + o(1)$  is *achievable*, while  $H(\mathcal{X}) - \Omega(1)$  is not. Here,  $o(1)$  denotes a positive quantity which can be chosen as small as we please, while  $\Omega(1)$  denotes a positive constant.

If two information source  $\mathcal{X}$  and  $\mathcal{Y}$  are correlated, then we have a  $\delta$ -joint distribution  $\mathcal{X}\mathcal{Y} = \{(x, y), p_{\mathcal{X}\mathcal{Y}}(x, y)\}$ . Therefore, the marginal distribution  $\mathcal{X}$  can be denoted as

$$\mathcal{X} = \{x, p_{\mathcal{X}}(x) = \sum_y p_{\mathcal{X}\mathcal{Y}}(x, y)\} \quad (2.6)$$

and similarly for  $\mathcal{Y}$ . A distribution is  $\delta$ -jointly typical if

$$\begin{aligned} 2^{-n(H(\mathcal{X})+\delta)} &\leq p(\vec{x}) \leq 2^{-n(H(\mathcal{X})-\delta)} \\ 2^{-n(H(\mathcal{Y})+\delta)} &\leq p(\vec{y}) \leq 2^{-n(H(\mathcal{Y})-\delta)} \\ 2^{-n(H(\mathcal{XY})+\delta)} &\leq p(\vec{x}, \vec{y}) \leq 2^{-n(H(\mathcal{XY})-\delta)} \end{aligned} \quad (2.7)$$

From the Bayes' rule we obtain the conditional probability  $p(x|y)$ , which defines the conditional entropy of  $\mathcal{X}$  given  $\mathcal{Y}$

$$H(\mathcal{X}|\mathcal{Y}) = H(\mathcal{XY}) - H(\mathcal{Y}) = - \sum_{x,y} p(x,y) \log_2 p(x|y) \quad (2.8)$$

This quantifies the remaining ignorance about  $x$  when  $y$  is known. This gap therefore needs  $H(\mathcal{X}|\mathcal{Y})$  bits to be bridged. Moreover, because  $\mathcal{XY}$  is a joint distribution, by observing one random variable, say  $y$ , we always know something of the other random variable  $x$ . The correlation between  $\mathcal{X}$  and  $\mathcal{Y}$  is thus quantified by the *mutual information*  $I(\mathcal{X}; \mathcal{Y})$ , which is symmetric and nonnegative (it is zero only if the distribution are completely uncorrelated). Nevertheless, the concept of mutual information is intimately linked to that of entropy of a random variable, in fact we have

$$I(\mathcal{X}; \mathcal{Y}) = H(\mathcal{X}) - H(\mathcal{X}|\mathcal{Y}) = H(\mathcal{Y}) - H(\mathcal{Y}|\mathcal{X}) = I(\mathcal{Y}; \mathcal{X}) \quad (2.9)$$

Therefore, because in a joint distribution situation the mutual information is always positive, holds that  $H(\mathcal{X}) \geq H(\mathcal{X}|\mathcal{Y}) \geq 0$ . Hence we notice that, by sampling over  $\mathcal{Y}$ , the ignorance about  $\mathcal{X}$ 's outcomes is reduced, and for that they can be further compressed.

### 2.1.1 Noisy channel coding

In classical information theory, to allow Alice to communicate with Bob, there occurs a channel through which encoded information has to pass. In real situations, this channel is also affected by transmission errors, making sometimes hard to reconstruct the original message. We start fixing an ensemble  $\mathcal{X} = \{x, p(x)\}$  for the input letters, and generate the codewords for a length- $n$  code with a certain *rate*  $R$  (*i.e.*, how many bits of error-free information are encoded in a physical bit transmitted through a channel subject to a certain level of random data transmission errors) and the code is known both by the sender, say, Alice and the receiver Bob. The upshot is that we are interested in optimizing this rate and, in order to do that, is necessary to generate the code in a suitable way. At this point, the *communication channel* transforms the code following certain rules (each channel is featured by specific actions it does on the code) producing an output message  $\vec{y} \in \mathcal{Y}$  that can be conveniently decoded, retrieving the original message. In this process some errors can occur with a specific channel error probability. Also some noise can corrupt the message causing errors. In this case, we say we have a *noisy channel* and we cannot

always decode with success a message sent over a noisy channel. However, Shannon stated that, as long as there is some correlation between the channel's input and output, we can convey an encoded  $nR$ -bit message without errors by sending one of the  $2^{nR}$   $n$ -letter codewords using the channel  $n \rightarrow \infty$  times. Where the rate  $R$  has an optimal nonzero value, that can be achieved, and corresponds to the mutual information  $I(\mathcal{X}; \mathcal{Y})$ . That is the *noisy channel coding theorem*.

Therefore, the channel is characterized by  $p(y|x)$ , the conditional probability that the letter  $y$  is received when the letter  $x$  is sent ([8]: section 10.1.4).

### Binary symmetric channel

In order to study how to achieve the optimal  $R$  value, we start considering a binary alphabet  $\{0, 1\}$  and the *binary symmetric channel*. In this type of communication each bit could be flipped with probability  $p$ , otherwise it can be received correctly with probability  $1 - p$ , so the conditional probabilities for each case are

$$\begin{aligned} p(0|0) &= 1 - p & p(0|1) &= p \\ p(1|0) &= p & p(1|1) &= 1 - p \end{aligned} \quad (2.10)$$

For each  $n$ , the code contains  $2^k$  *codewords* among the  $2^n$  possible strings of length  $n$ , and the rate  $R$  is defined as

$$R = \frac{k}{n} \quad (2.11)$$

For any  $n$ -bit input message, we expect about  $np$  bits to be flipped, in fact selected and sent one of the  $2^{nR}$  codewords, it diffuses into one of about  $2^{nH(p)}$  typical output strings, occupying an error sphere of *Hamming radius*  $np$  and center in the input string [8]. To protect against errors, we should use a code such as its codewords are separable, that means the *Hamming distance* between two codewords is enough to permit the error spheres of two different codewords to not overlap. Otherwise, two different inputs will sometimes yield the same output, causing decoding errors to occur. A necessary condition for realizing this is to have no more than the total number of output message bits for the total number of strings contained in all of the  $2^{nR}$  spheres

$$2^{nH(p)} 2^{nR} \leq 2^n \quad (2.12)$$

which implies

$$R \leq 1 - H(p) = C(p) \quad (2.13)$$

Therefore,  $C(p)$ , which is the *channel capacity*, is the optimal rate for reliable transmission over the noisy channel. Shannon's most ingenious idea was that the rate  $C(p)$  can be achieved by an average over *random codes*. The easiest way to perform this is to generate the  $2^{nR}$  random codewords by sampling a total of  $2^{nR}$  times from the distribution  $\mathcal{X}^n$  (i.e., sampling over  $\mathcal{X}$   $n$  times). To send  $nR$  bits of information, Alice chooses one of the codewords and sends it to Bob by using the channel  $n$  times. In order to decode the message, Bob draws an Hamming sphere with a radius slightly large than  $np$  that contains  $2^{n(H(p)+\delta)}$  strings. If the sphere contains a

unique codeword, Bob decodes accordingly, otherwise he decodes arbitrarily. Since the sphere is large enough to contain with high probability, in an asymptotically regime, the codeword sent by Alice, only remains to worry about the sphere might accidentally contain another codeword. Another time, we find that the probability that the sphere contains one of the  $2^{nR} - 1$  invalid codewords is no more than

$$2^{nR} \frac{2^{n(H(p)+\delta)}}{2^n} = 2^{-n(C(p)-R-\delta)} \quad (2.14)$$

Since  $\delta$  may be as small as we please, we can choose  $R = C(p) + c$ , with  $c$  a positive constant, and this probability goes to zero as  $n \rightarrow \infty$ . At this point, the proof works for a fixed message. But we can assure this probability goes to zero asymptotically for every codeword ([8]: section 10.1.4).

### Channel capacity

Now consider to apply this random coding argument to more general alphabets and channels. The input ensemble  $\mathcal{X}$ , together with the conditional probability characterizing the channel, determines the joint ensemble  $\mathcal{X}\mathcal{Y}$  for each letter sent, due to the identity  $p(x, y) = p(y|x)p(x)$ . In order to reconstruct the message received by Bob, one has to use the notion of joint typicality in (2.7). When Bob receives an  $n$ -letter message  $\vec{y}$ , he determines whether there exists an  $n$ -letter initial message  $\vec{x}$ , so that  $\vec{x}$  is jointly typical with  $\vec{y}$ . If this happens, Bob decodes accordingly, otherwise he decodes arbitrarily.

In order to compute the probability of a decoding error to occur, we notice that by applying the strong law of large numbers to (2.7), the probability for a  $(\vec{x}, \vec{y})$  of being jointly  $\delta$ -typical is given by (2.5). Therefore, it only remains to ensure that the probability  $p(\vec{x}', \vec{y})$  of any other codeword  $\vec{x}'$  to be *jointly*  $\delta$ -typical with  $\vec{y}$  will vanish asymptotically (note that every  $(\vec{x}', \vec{y})$  is  $\delta$ -typical with probability (2.5)). From (2.7), we find that the number  $N_j$  of jointly  $\delta$ -typical  $(\vec{x}, \vec{y})$  is

$$N_j \leq 2^{n(H(\mathcal{X}\mathcal{Y})+\delta)} \quad (2.15)$$

Moreover, because the samples are independent, the probability of drawing two codewords factorizes as  $p(\vec{x}', \vec{x}) = p(\vec{x}')p(\vec{x})$ . Consequently, the channel output  $\vec{y}$ , when  $\vec{x}$  is sent, is independent from  $\vec{x}'$  and  $p(\vec{x}', \vec{y}) = p(\vec{x}')p(\vec{y})$ . Therefore, from (2.7), we have

$$\sum_{\text{jointly } (\vec{x}', \vec{y}) \text{ couples}} p(\vec{x}')p(\vec{y}) \leq N_j 2^{-n(H(\mathcal{X})-\delta)} 2^{-n(H(\mathcal{Y})-\delta)} \leq 2^{-n(I(\mathcal{X};\mathcal{Y})-3\delta)} \quad (2.16)$$

Finally, considering  $2^{nR}$  codewords independently generated by sampling  $n$ -times from  $\mathcal{X}$ , the probability became

$$2^{nR} 2^{-n(I(\mathcal{X};\mathcal{Y})-3\delta)} = 2^{n(R-I(\mathcal{X};\mathcal{Y})+3\delta)} \quad (2.17)$$

Since we can choose  $\delta$  as small as we please, we can write  $R = I(\mathcal{X}; \mathcal{Y}) - c$ , with  $c > 0$  is a constant. Therefore we have demonstrated that this probability approach to zero asymptotically ([8]: section 10.1.4). That induces the definition of channel capacity

$$C = \sup_{\mathcal{X}} I(\mathcal{X}; \mathcal{Y}) \quad (2.18)$$

This allows to interpret the mutual information as the information per letter we can transmit over the channel. In this sense the mutual information quantifies the information gained about  $\mathcal{X}$  when we have access to  $\mathcal{Y}$ . Notice that we have obtained a formula for the capacity just for a *single use* of the channel, although the capacity is achieved by many letters messages.

Ultimately, the demonstration Shannon provides of  $R = I(\mathcal{X}; \mathcal{Y}) - c$  being reachable averaging over random codes, is non constructive. Since a random code has no structure or pattern, to encode and decode requires an exponentially large code book. It is very interesting and useful to look for codes which can be efficiently encoded and decoded, and come close to achieve the capacity.

## 2.2 Von Neumann Entropy

We start now to generalize the considerations above to quantum information. Firstly, imagine that the letter of a message are picked up from an ensemble of quantum states  $\{\sigma(\mathbf{x})\}$ , each occurring with a specific a priori probability  $p(x)$ . As said in Chapter 1, all the probabilities of the measurement outcomes without knowing nothing about the system, are specified by the density operator  $\rho = \sum_x p(x)\sigma(\mathbf{x})$ . For a *POVM* the probability of outcome  $a_i$  is given by (1.9). Therefore, for any density operator, we define the Von Neumann entropy as:

$$H(\rho) = -\text{tr}(\rho \log \rho). \quad (2.19)$$

Now we consider a basis  $\{a_i\}$  that diagonalizes  $\rho$  as in (1.6). Hence, the vector of eigenvalues of the density operator  $\lambda(\rho)$  is a probability distribution and, with simple calculations, the Von Neumann entropy of  $\rho$  is just the Shannon entropy of this distribution

$$H(\rho) = H(\lambda(\rho)) \quad (2.20)$$

The central issue in quantum information theory is that states are nonorthogonal in general and therefore they cannot be perfectly distinguished. But, drawing from an alphabet of mutually orthogonal and therefore completely distinguishable pure quantum states  $\{|\varphi(x)\rangle, p(x)\}$ , an  $n$ -letter message can be compressed, without decoding errors, to a one of  $H(\rho)$  qubits per letter. In fact, the density operator of this ensemble  $\rho = \sum_x p(x) |\varphi(x)\rangle \langle \varphi(x)|$  has the property to be  $\rho^{\otimes n} = \rho \otimes \dots \otimes \rho$  and it turns out that the message can be compressed to a Hilbert space so that  $\dim \mathcal{H} = 2^{n(H(\rho)+o(1))}$  asymptotically ([8]: sections 10.3). That is the *Quantum Source Coding Theorem*. Moreover, in this case also holds that  $H(\rho) = H(\mathcal{X})$ ,

causing the Von Neumann entropy to quantify the maximal *classical* information per letter gained by making the best possible measurement. Finally, since for a pure state the density matrix is idempotent, the entropy  $H(\boldsymbol{\rho})$  vanishes, so it gives us the departure of a system from being a pure state by quantifying its entanglement ([8]: section 10.2). In fact, considering the case where  $\boldsymbol{\rho}$  has  $d$  non-vanishing eigenvalues, holds that  $H(\boldsymbol{\rho}) \leq \log d$ , with equality when all the nonzero eigenvalues are equal. This causes for the entropy of a maximally mixed state to be maximized. Moreover, by investigating more deeply, if  $\boldsymbol{\rho}_{AB}$  is a bipartite pure state, then  $H(A) = H(B)$  as  $\boldsymbol{\rho}_A$  and  $\boldsymbol{\rho}_B$  share the same eigenvalues.

Nevertheless, a natural definition of quantum mutual information comes from (2.9) by substituting the Von Neumann entropy.

At this point, is useful to introduce the *relative entropy*. In a classical context, the relative entropy of a probability distribution  $\{p(x)\}$  relative to  $\{q(x)\}$  is defined as

$$D(p\|q) \equiv \sum_x p(x)(\log p(x) - \log q(x)) \quad (2.21)$$

and one can easily demonstrate (using the disequation  $\log x \leq x - 1$ , for  $x$  positive and real, where we have equality only if  $x = 1$ ) that  $D(p\|q) \geq 0$ , where equality holds if and only if the probability distributions are identical. Therefore, we accordingly define the quantum relative entropy of  $\boldsymbol{\rho}$  with respect to  $\boldsymbol{\sigma}$  as

$$D(\boldsymbol{\rho}\|\boldsymbol{\sigma}) = \text{tr } \boldsymbol{\rho}(\log \boldsymbol{\rho} - \log \boldsymbol{\sigma}) \quad (2.22)$$

which, denoting  $\{p_i\}$  as the eigenvalues of  $\boldsymbol{\rho}$  and  $\{q_i\}$  as the ones of  $\boldsymbol{\sigma}$ , operatively becomes

$$D(\boldsymbol{\rho}\|\boldsymbol{\sigma}) = \sum_i p_i \left( \log p_i - \sum_a D_{ia} \log q_a \right) \quad (2.23)$$

Here,  $D_{ia}$  is a double stochastic matrix (its entries are nonnegative real numbers and each column or row is a probability distribution). With a little extra effort (an approach to the proof is given in [8]: exercise 10.1), one could find that

$$D(\boldsymbol{\rho}\|\boldsymbol{\sigma}) \geq 0 \quad (2.24)$$

with equality if and only if  $\boldsymbol{\rho} = \boldsymbol{\sigma}$ . This property is known as the *positivity* of quantum relative entropy. Another important property is called *monotonicity*:

$$D(\boldsymbol{\rho}_A\|\boldsymbol{\sigma}_A) \leq D(\boldsymbol{\rho}_{AB}\|\boldsymbol{\sigma}_{AB}) \quad (2.25)$$

### 2.2.1 Entropy and thermodynamics

The concept of entropy first entered science through the study of thermodynamics and is important to see how quantum information theory can illuminates it. Moreover, nonnegativity and monotonicity of quantum relative entropy strongly relates

to ideas in thermodynamics.

First of all, we consider an approach where the evolution of the full system is unitary, but not the one of the subsystem is not, and it can be accurately described by a thermal ensemble at late times. Therefore, if information is initially encoded locally into a non-equilibrium state, becomes more and more non local as the system evolves, eventually becoming invisible to an observer on the subsystem. For this reasons, the state of an open system with Hamiltonian  $\mathbf{H}$  is expected to be close to the thermal Gibbs state

$$\rho_B = \frac{e^{-\beta\mathbf{H}}}{\text{tr}(e^{-\beta\mathbf{H}})} \quad (2.26)$$

where  $kT = \beta^{-1}$  is the temperature. For an arbitrary  $\rho$  we define the free energy as

$$F(\rho) = E(\rho) - \beta^{-1}S(\rho) \quad (2.27)$$

and  $E(\rho) = \langle \mathbf{H} \rangle_\rho$  is the expectation value of the Hamiltonian in this state. Using the positivity of relative entropy, we can easily demonstrate that, at a temperature  $\beta^{-1}$ , the Gibbs state has the lowest possible free energy. In fact, we have

$$\begin{aligned} F(\rho) &= \text{tr}(\rho\mathbf{H}) - \beta^{-1}H(\rho) = \beta^{-1} \text{tr} \rho(\log \rho + \beta\mathbf{H}) \\ F(\rho_\beta) &= -\beta^{-1} \log(\text{tr} e^{-\beta\mathbf{H}}) = \beta^{-1} \text{tr}(\beta\mathbf{H}) \end{aligned} \quad (2.28)$$

and therefore the relative entropy of  $\rho$  and  $\rho_\beta$  is

$$D(\rho\|\rho_\beta) = \beta(F(\rho) - F(\rho_\beta)) \geq 0 \quad (2.29)$$

Moreover, using monotonicity of relative entropy and the fact that the joint unitary evolution of the system induces a quantum channel  $\mathcal{N}$  acting on the system alone, descends that a quantum channel cannot increase relative entropy  $D(\mathcal{N}(\rho)\|\mathcal{N}(\rho_\beta)) \leq D(\rho\|\rho_\beta)$ . Furthermore, we also expect that  $\mathcal{N}$  preserves the Gibbs equilibrium state. This yields an alternative version of the second law of thermodynamics ([8]: section 10.2.5)

$$D(\mathcal{N}(\rho)\|\mathcal{N}(\rho_\beta)) = \beta(F(\mathcal{N}(\rho)) - F(\rho_\beta)) \leq \beta(F(\rho) - F(\rho_\beta)) = D(\rho\|\rho_\beta) \quad (2.30)$$

and hence

$$F(\mathcal{N}(\rho)) \leq F(\rho) \quad (2.31)$$

Thus, the free energy of a non-equilibrium state is monotonically decreasing under open-state evolution.

### 2.2.2 Bekenstein's entropy bound

A very interesting application for what said above can be found in quantum field theory formulation of Bekenstein's bound on entropy. More specifically, the bound is formulated as an inequality relating the energy and the entropy in a bounded



spatial region. The idea Bekenstein's had of such a bound was motivated by issues about black hole thermodynamics and gravitational physics, but it can be formulated without reference to gravitation (we briefly see why), and follows from properties of relative entropy. However, a sketch of the black hole thermodynamics and the original formulation of this bound are central points in the Appendix A of this work. We start considering a two-dimensional plane in which we draw an edge, defining a region. According to quantum field theory, in the vacuum state virtual processes can occur. That is, couples of particle and antiparticle are created in vacuum and they annihilate in a time so short that, because of  $\Delta E \Delta t \sim \hbar$ , a violation of energy-momentum relation can occur. Because that continuously happens, there are infinite contributes to the 'vacuum energy' guaranteed by these processes, causing for the entropy of a region to be infinite in quantum field theory. Therefore, if we want to define a suitable 'vacuum state' being a finite energy state, quantum field theory offers us the possibility to subtract this background energy caused by virtual processes. Operatively, this corresponds to a rescaling of the Hamiltonian  $\mathbf{H}$ , which becomes  $\mathbf{K}$  after setting the temperature to unity. Hence, although the vacuum is a pure state, when we consider only a region of a plane its marginal state becomes highly mixed because of the entanglement between it and its complement. So, the vacuum mixed state is

$$\rho_0 = \frac{e^{-\mathbf{K}}}{\text{tr}(e^{-\mathbf{K}})} \quad (2.32)$$

Then, for any state  $\rho$ , the positivity of relative entropy yields

$$H(\rho) - H(\rho_0) \leq \text{tr}(\rho \mathbf{K}) - \text{tr}(\rho_0 \mathbf{K}) \quad (2.33)$$

Therefore the difference between the entropy of a given state of the region and its 'vacuum entropy' is bounded above by the same difference between (modular) energies. This is a form of Bekenstein's bound. What is noteworthy is that  $\mathbf{K}$  is dimensionless and extensive, therefore (in units with  $\hbar = c = 1$ ) can be interpreted as  $ER$ , where  $R$  is the linear size of the region. The justification of this fact involves relativistic quantum field theory and is beyond the aim of this work and is not given here. The upshot is that the entropy of a two-dimensional region is bounded by  $O(ER)$  and therefore by the edge of the region. This statement, with a dimension more, yield the other version of Bekenstein's bound given for black holes which involves the area  $A$  of the event horizon. The formula is (A.1), and it is widely justified and discussed, together with the whole Bekenstein's bound in 'black holes language', in the section A.1.1. Another important consideration that can be made on (2.33), is that, considering the Hawking Radiation process, the right-hand side can be negative. In fact, when couples of virtual particles are created in the correspondence of the event horizon and the particle with energy  $E$  runs out of the gravitational field of the black hole by tunnelling, the other one, with energy  $-E$ , falls into the black hole by decreasing its total energy. Therefore, if we perform the energy subtraction we obtain a negative result. This implies that the Hawking's phenomenon causes entropy of a region to decrease, and this is possible (remember

---

the connection between entropy of a region and its boundary) only by reducing its surface. Therefore, black holes actually evaporates and decrease their size with time. More details in A.1.2, where also a better description of Hawking radiation is given.

## Chapter 3

# Quantum Channel Capacities and Decoupling

At this point of the work, we want to deepen some general aspects related to quantum channels. The concept of decoupling will also be a main theme in this chapter and we will see how it can be applied in the study of an informative process considering an idealized model of black hole dynamics. The application of quantum information we will provide will be fundamental in order to understand some complex aspects of this theory and relate black holes physics to quantum information theory.

First of all, we want to focus a little more on a property of Von Neumann entropy. For a bipartite system holds the triangle inequality ([8]: section 10.2.1)

$$H(AB) \geq |H(A) - H(B)| \quad (3.1)$$

which strongly contrast with the classical equivalent  $H(AB) \geq H(A), H(B)$ . This causes that in a bipartite pure system, one has  $H(A) = H(B) > 0$  and therefore  $H(AB) = 0$ . Conversely, this situation can classically occur only if  $H(A), H(B) = 0$ . Hence, in the definition of Von Neumann conditional entropy, if  $\rho_{AB}$  is an entangled bipartite pure state, one could have that

$$H(A|B) = H(AB) - H(B) = -H(B) < 0 \quad (3.2)$$

This happens due to the stronger-than-classical correlation force for entangled pure quantum states. The negative of the conditional quantum entropy is so important in quantum information theory that it even has a special name: the *coherent information*. For  $\rho_{AB}$ , the coherent information from  $A$  to  $B$  is as follows

$$I(A)B)_\rho \equiv H(B)_\rho - H(AB)_\rho \quad (3.3)$$

### 3.1 Quantum Capacity Theorem and Decoupling approach

We want now to find a regularized formula for the quantum capacity of a quantum channel  $Q(\mathcal{N}^{A \rightarrow B})$ . Firstly, remember that a quantum channel  $\mathcal{N}^{A \rightarrow B}$  is a TPCP map from the space  $\mathcal{H}_A$  to  $\mathcal{H}_B$ . Alice uses the channel  $n$  times for sharing a quantum state with Bob. She prepares  $\psi$  in a code subspace  $\mathcal{H}^{(n)} \subseteq \mathcal{H}_A^{\otimes n}$  and sends it to Bob, which applies a decoder in order to recover  $\psi$ . Remember also that the rate is the number of qubits sent per channel use, and basically it is achievable if there exists a sequence of codes with rate at least  $R - \delta$  and Bob's state  $\rho$  has a fidelity  $\langle \psi | \rho | \psi \rangle \geq 1 - \varepsilon$ , for every  $\varepsilon, \delta > 0$ . Moreover, recall that any channel  $\mathcal{N}^{A \rightarrow B}$  has an isometric *Stinespring* dilatation  $\mathbf{U}^{A \rightarrow BE}$ , where  $E$  is the channel environment ([6]: section 3.3.2). Furthermore, suppose now the input code state is  $\rho_A$ . In order to deliver a pure state through the channel, it has a purification by introducing a reference system  $R$ , such that  $\rho_A = \text{tr}(|\psi\rangle\langle\psi|)$ . Applying the channel's dilatation to  $\psi_{RA}$ , we obtain an output state  $\phi_{RBE}$ . Therefore, for the class of *degradable channels* ([6]: section 3.2), the *one-shot* coherent information is our best characterization for quantum capacity ([8]: section 10.7.1)

$$Q(\mathcal{N}) = \max_A (I(R)B)_{\phi_{RBE}} \quad (3.4)$$

where the maximum is taken over the all possible input density operator  $\{\rho\}$ .

An approach to proving the quantum capacity formula (3.4), is the decoupling approach [14]. Suppose that  $\rho$  is a quantum code state and Alice wants to share its purification  $\psi_{RA}$  in the reference system with Bob. In order to do that she ships her state through the dilatation  $\mathbf{U}^{A \rightarrow BE}$ . As a result, we have a tripartite pure entangled state  $\phi_{RBE}$ . Then, if the reduced state  $\psi_{RE}$  on the reference system and environment system is approximately *decoupled* (*i.e.*, is almost a product state), meaning that

$$\|\psi_{RE} - \psi_R \otimes \sigma_E\|_1 \leq \varepsilon \quad (3.5)$$

where  $\sigma_E$  is some arbitrary state, we found that Bob is able to recover perfectly the action of the channel dilatation  $\mathbf{U}^{A \rightarrow BE}$  on the pure state  $\psi_{RA}$ . (In (3.5), the trace norm  $\|X\|_1$  of an operator  $X$  is the sum of its singular values). In order to understand why this happens, let us suppose that the state is *exactly* decoupled (exact correctability corresponds to exact decoupling, but likewise approximate correctability corresponds to approximate decoupling ([8]: section 10.7.2)). Then, one purification of the state  $\psi_{RE}$  is the state  $\phi_{RBE}$  that results after the channel acts. Since all purifications are equal up to a unitary, another purification for  $\psi_{RE} = \psi_R \otimes \sigma_E$  is

$$\tilde{\psi}_{RB_1} \otimes \sigma_{B_2E} \quad (3.6)$$

where  $B$  decomposes into  $B = B_1B_2$ . Moreover,  $\tilde{\psi}_{RB_1}$  is the original state that Alice sends through the channel and  $\sigma_{B_2E}$  is some other state that purifies the state

$\sigma_E$  of the environment. Since all purifications are related by isometries, and since Bob holds in  $B_1$  the purification of the state of  $R$  and in  $B_2$  the one of  $E$ , there exists some unitary  $U^{B \rightarrow B_1 B_2}$  such that

$$U^{B \rightarrow B_1 B_2} |\psi\rangle_{RBE} = |\psi\rangle_{RB_1} \otimes |\sigma\rangle_{B_2 E} \quad (3.7)$$

Then, this unitary is Bob's decoder. Thus, the decoupling condition implies the existence of a decoder for Bob. Moreover, by admitting that Bob holds in  $B_1$  the purification of the state of  $R$ , a consequence of the decoupling condition is that

$$H(R) = I(R)A)_{\psi} = I(R)B_1)_{\tilde{\psi}} \quad (3.8)$$

Therefore, we have demonstrated, in the case of perfect decoupling, that the coherent information that Bob receives and is able to decode is exactly the same that Alice transmits. And therefore, the number of qubits sent per channel use is exactly the coherent information from  $R$  to  $B$ . That proves (3.4) in this special situation.

We may chose the initial state to be a maximally entangled state  $\phi_{RA}$ . Then, if the resulting  $\phi_{RBE}$  has a marginal state  $RE$  which factorizes, then, by the *relative state method* in 1.2.7, we conclude that any state in the code space can be sent to Bob and decoded with perfect fidelity.

To conclude, we found that purified quantum information sent trough a noisy channel is exactly correctable if and only if the reference system is completely uncorrelated with the channel's environment. That is the *decoupling principle*.

## 3.2 The Decoupling Inequality

We proved that (3.4) is an upper bound on the capacity by using the decoupling approach. In particular, we argued that it is sufficient to design codes which remain decoupled from the environment in order to perfectly recover the initial message sent trough the channel, and we said that only exact correctability corresponds to exact decoupling, but also approximate correctability works with approximate decoupling. Moreover, we found that this machine works accordingly to (3.5), that can be considered as a sufficient condition of decoupling from environment.

At this point, we introduce that, as for classical Shannon theory, achievable rates for quantum protocols are derived by using random codes ([8]: section 10.8). However, this similarity is superficial and the condition of decoupling under this condition needs to be tailored by introducing a way to decouple a uniformly random subspace of a subspace of the input. Following this idea, we reach the *decoupling inequality* or the *One-shot decoupling theorem*. Let us try to clarify this point by making an example: suppose that Alice has a quantum state  $\sigma_{AE}$  in sharing with Eve's system  $E$ , where  $A$  in an  $n$ -qubit system, correlated with  $E$ , so that  $I(A; E) > 0$ . Now, we wonder how is the amount of qubits that Alice has to discard so that the subsystem she retains has a negligible correlation with  $E$ . At this point is necessary to formalize what it means to discard a random qubit. Suppose that  $A$  has a dimension  $|A|$ , and

$A$  is decomposed into two subsystems  $A_1$  and  $A_2$ , then discards  $A_1$  and retains  $A_2$ . Therefore, discarding a random subsystem with dimension  $A_1$  is the same thing as applying a random unitary  $\mathbf{U}$  before discarding the fixed subsystem  $A_1$ . More in detail, to choose a random subsystem to discard is the same of averaging  $\mathbf{U}$  uniformly over the group of unitary  $|A| \times |A|$  matrices. We denote the expectation value of a function  $f(\mathbf{U})$  in this conditions, as  $\mathbb{E}[f(\mathbf{U})]$ . Alternatively, when unitaries  $\mathbf{U}$  are uniformly distributed, we describe  $\mathbb{E}[f(\mathbf{U})]$  as the integral over the unitary group using the *Haar measure* on the group ([8]: section 10.9). When  $\mathbf{U}$  is applied to  $A$ , and then discard  $A_1$ , the marginal state of  $A_2E$  is

$$\sigma_{A_2E}(\mathbf{U}) = \text{tr}_{A_1} \left( (\mathbf{U}_A \otimes \mathbf{I}_E) \sigma_{AE} (\mathbf{U}_A^\dagger \otimes \mathbf{I}_E) \right) \quad (3.9)$$

Therefore, we reach the decoupling inequality that express how close  $\sigma_{A_2E}$  is to a product state when we average over  $\mathbf{U}$ . Hence, it rewrites (3.5)

$$\mathbb{E} [\|\sigma_{A_2E}(\mathbf{U}) - \sigma_{A_2}^{\max} \otimes \sigma_E\|_1] \leq \sqrt{\frac{|A_2| \cdot |E|}{|A_1|} \text{tr}(\sigma_{AE}^2)} \quad (3.10)$$

where  $\sigma_{A_2}^{\max} = I/|A_2|$  is the maximally mixed state on  $A_2$ . A complete proof of this inequality can be found in [14] or alternatively in [8]: section 10.9.1. Anyway, to give the main ideas of the demonstration, we start considering the Cauchy-Schwartz inequality and the concavity of the square-root function

$$\begin{aligned} \mathbb{E} [\|\sigma_{A_2E}(\mathbf{U}) - \sigma_{A_2}^{\max} \otimes \sigma_E\|_1] &\leq \mathbb{E} \left[ \sqrt{|A_2||E| \|\sigma_{A_2E}(\mathbf{U}) - \sigma_{A_2}^{\max} \otimes \sigma_E\|_2^2} \right] \\ &\leq \sqrt{|A_2||E| \text{Var} [\sigma_{A_2E}]} \end{aligned} \quad (3.11)$$

where in the last line we use the variance of  $\sigma_{A_2E}$

$$\begin{aligned} \text{Var} [\sigma_{A_2E}] &= \mathbb{E} [\|\sigma_{A_2E}(\mathbf{U}) - \sigma_{A_2}^{\max} \otimes \sigma_E\|_2^2] \\ &= \mathbb{E} [\text{tr}(\sigma_{A_2E}^2)] - \frac{1}{|A_2|} \text{tr}(\sigma_E^2) \end{aligned} \quad (3.12)$$

Therefore, it only suffices to prove that

$$\mathbb{E} [\text{tr}(\sigma_{A_2E}^2)] \leq \frac{1}{|A_2|} \text{tr}(\sigma_E^2) + \frac{1}{|A_1|} \text{tr}(\sigma_{AE}^2) \quad (3.13)$$

in order to prove (3.10) as desired. In [8] and [14] we see this can be performed in two slightly different ways.

To conclude, we show that, provided the dimension of the discarded part  $|A_1|$  is sufficiently small with respect to the retained part  $|A_2|$ , the transmitted data will with high probability be decoupled from the channel's environment. Moreover, all what said yields an inequality that specifies a sufficient condition for decoupling when the average input is close to a product state. From this, it is possible to

compute the quantum capacity of an arbitrary quantum channel as far as the code randomness is expressed by the randomness with which Alice selects and discards a qubit from  $A_2$  (forming  $A_1$ ) in order to decouple it from Eve's one.

The last thing to highlight is that, from (3.10), by randomly choosing a pure state of the bipartite system  $A = A_1 A_2$ , where  $|A_1|/|A_2| \ll 1$  we obtain codes (all the possible  $n$ -qubit systems of dimension  $|A_2|$ ) which are maximally entangled with a uniformly random subspace  $|A_1|$  of  $A$ . Thus, the code system, full of information, and the discarded qubit system are maximally entangled with high probability under certain conditions regarding their dimensions according to a slightly modified version of (3.10). This will be fundamental for results obtained in the next paragraph.

### 3.3 Black holes as mirrors

At the end of the work, we consider an application of the decoupling inequality in a highly idealized model of black hole dynamics ([9]; [8]: section 10.9.5). Suppose Alice holds a  $k$ -qubit system  $A$  which she wants to conceal from Bob. She decides to discard her qubits by throwing them into a large black hole  $B$ , that is an  $(n - k)$ -qubit system, which grows to  $n$  qubits after merging with  $A$ , where  $n$  is much larger than  $k$ . Black holes are not completely black, in fact they emit Hawking radiation (see Appendix section A.1.2). Qubits leak out of an evaporating black hole very slowly, at a rate given by (A.11) which scales like  $n^{-1/2}$ . Consequently, we notice that from (A.13) descends that the black hole radiates away a significant part of its qubits after a time which is  $O(n^{3/2})$ . Therefore, for black holes of a mass close to the solar mass, the evaporation process takes about  $10^{68}$  years to complete (A.1.2). Although Alice's qubits might not remain secret forever, they will be safe from Bob for at least a so long time. Unfortunately for Alice, in this case we consider a very old black hole. In fact, it has been evaporating for so long that it has already radiated away more than half of its qubits. We assume that the internal dynamics of the black hole is a deterministic unitary transformation that accurately mixes the infalling information with the black hole's preexisting  $(n - k)$ -qubit state. Then, the black hole's qubits are released, one by one, in the Hawking radiation. In the whole process is fundamental to consider how the entanglement of the black hole with the emitted radiation evolves.

The world can be divided into two subsystems: the black hole's internal system  $B$  and radiated system  $E$ . The relative size of these subsystems varies with time as the black hole evaporates. Furthermore, let us assume that the joint state of the black hole  $B$  and its emitted radiation  $E$  is *pure*. Because the black hole  $B$  is so old,  $|B|$  is much smaller than the dimension of the radiation system and, for what we saw in the previous section 3.2, we expect the state of  $B$  to be very nearly maximally mixed with high probability. Therefore, we claim that  $B$  and  $E$  are very nearly maximally entangled. Consider also the system  $A$  to be maximally entangled with a reference system  $R_A$  which also consists of a purification for  $A$ , causing  $R_A A$  to be pure. We also assume that  $|R_A| = |A|$ . Right after Alice tosses in her qubits, the

$n$ -qubit black hole system  $AB$  (remind that is impossible to distinguish individual parts of  $A$  from the ones of  $B$  in  $AB$ ) is maximally entangled with the system  $R_A E$ ; here  $E$  is the previously emitted Hawking radiation and we assume that it has been collected and now controlled by Bob.

We wonder if Bob would be capable of recovering the initial message Alice tossed in black holes, and we claim that, for almost any unitary transformation, Bob needs to wait for only a few more than  $k$  qubits to be emitted to reach the scope. In fact, as Bob noticed, the black hole continues to emit Hawking radiation until, after a while,  $s$  additional qubits (the subsystem  $\tilde{A}$  of  $AB$ ) have been emitted, with  $n - s$  qubits (the subsystem  $\tilde{AB}$ ) still retained by the black hole. It is important to specify that we suppose that  $\tilde{A}$  is chosen uniformly at random (is an *Haar random* subsystem of the whole system), and that  $|\tilde{A}| > |A|$ . That is, we imagine that  $AB$  is divided into two parts, one  $s$ -qubit part and another with  $n - s$  qubits. Then, a unitary transformation  $V^{AB}$  chosen uniformly with respect to the Haar measure on the group of unitaries acting on  $n$ -qubits strings  $U(2^n)$  is applied to  $AB$ . After this operation, the  $s$ -qubit system is identified as  $\tilde{A}$ .

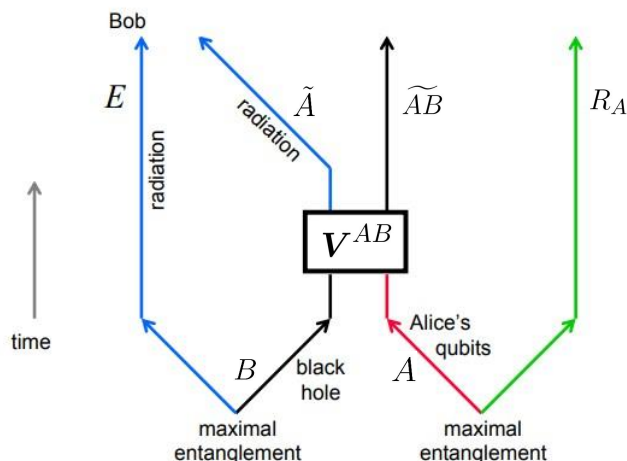


Figure 3.1: Image elaborated from the one in [9], illustrating the process of information recovery from black holes evaporation. First of all, because the black hole is very old and the radiative process is going on for so long,  $|B| \ll |E|$  and they are maximally mixed. System  $A$ , maximally entangled with  $R_A$ , is thrown into  $B$ , mixed up and later transformed by  $V^{AB}$ . Bob waits until a system  $|\tilde{A}|$ , slightly greater than  $|A|$ , is discarded by emitted radiation. With  $E\tilde{A}$  nearly maximally entangled with  $R_A$ , Bob holds Alice initial information on his hand.

As the radiative process goes on, the correlations between the evaporating black hole  $\tilde{AB}$  and the reference system  $R_A$  gradually weaken, therefore we might say they are decoupling. Once  $\tilde{A}$  is large enough, the surviving correlation of  $R_A$  with



$\widetilde{AB}$  becomes negligible, according to (3.10). At that point, since the overall state of  $\widetilde{ABR}_A\widetilde{AE}$  is pure, the state of the reference system  $R_A$  is very nearly purified by the radiation system  $\widetilde{AE}$  that Bob controls, as depicted in figure 3.1. More precisely, Bob can succeed in recovering the purification of  $R_A$  by applying a suitable decoding map to a subsystem of  $E\widetilde{A}$ . This because of  $|\widetilde{A}| > |A|$  and the decoupling inequality. In other words, Alice's quantum information is now property of Bob.

Because the radiated qubits are random, we can determine the conditions of decoupling to occur between  $R_A$  and  $\widetilde{AB}$  using the decoupling inequality. Let  $\psi^{ABR_AE}$  be the pure density operator of  $ABR_AE$ . We denote with  $\sigma_{\widetilde{ABR}_A}$  the marginal state  $\sigma_{\widetilde{ABR}_A} = \text{tr}_E(\psi^{ABR_AE})$ . The marginal density operator on  $\widetilde{ABR}_A$  is

$$\sigma_{\widetilde{ABR}_A}(\mathbf{V}^{AB}) = \text{tr} [\sigma_{ABR_A}(\mathbf{V}^{AB})], \text{ where } \sigma_{ABR_A}(\mathbf{V}^{AB}) \text{ is given by (3.9).}$$

The inequality so becomes

$$\begin{aligned} \mathbb{E} \left[ \|\sigma_{\widetilde{ABR}_A}(\mathbf{V}^{AB}) - \sigma_{\widetilde{AB}}^{\max} \otimes \sigma_{R_A}\|_1 \right] &\leq \sqrt{\frac{|\widetilde{AB}| \cdot |R_A|}{|\widetilde{A}|} \text{tr}(\sigma_{\widetilde{ABR}_A}^2)} \\ &\leq \sqrt{\frac{|AB| \cdot |R_A|}{|\widetilde{A}|^2} \text{tr}(\sigma_{\widetilde{ABR}_A}^2)} \end{aligned} \quad (3.14)$$

where we used the fact that  $|\widetilde{AB}| = |AB|/|\widetilde{A}|$ . Moreover, in the case we are currently considering,  $AB$  is maximally entangled with  $R_AE$  causing for  $\sigma_{\widetilde{ABR}_A}$  to be maximally mixed on a system of dimension  $|AB|/|R_A|$ . Therefore we have

$$\text{tr}(\sigma_{\widetilde{ABR}_A}^2) = |R_A|/|AB| \quad (3.15)$$

hence the Haar-averaged  $L^1$  distance of  $\sigma_{\widetilde{ABR}_A}$  from a product state is bounded above by

$$\sqrt{\frac{|R_A|^2}{|\widetilde{A}|^2}} = \frac{|A|}{|\widetilde{A}|} \quad (3.16)$$

where we used that  $|R_A| = |A|$ . Thus, if Bob waits for only  $s = k + c$  qubits of Hawking radiation to be emitted after  $k$ -qubit system  $A$  getting tossed into black hole, Bob can decode with fidelity  $F \geq 1 - |A|/|\widetilde{A}| = 1 - 2^k/2^s = 1 - 2^c$ .

Therefore, in order to decode Alice's secret, Bob needs only  $s$  qubits more after Alice's attempt to conceal  $A$  from him. And Bob, who is an excellent physicist, knows the black hole thermodynamics enough to infer the right encoding unitary  $\mathbf{V}^{AB}$ , that he uses to find the decoding map.

So far we assumed that the system  $\widetilde{A}$  is a randomly selected subsystem of  $AB$ . For a real black hole this is incorrect, but things are believed to be similar to this approximation thanks to the internal dynamics that mixes quantum information quite rapidly relying on a process known as *fast scrambling procedure*. We have treated in greater detail this argument on the section A.1.3 of the Appendix, where

we also specified that for a black hole of temperature  $T$ , it takes a time  $\hbar/kT$  for each qubit to be emitted in the Hawking radiation, and a time longer by only a factor of  $\log(n)$  for the dynamics to mix the degrees of freedom sufficiently for decoupling to hold with reasonable accuracy. Follows that, for a solar mass black hole, the qubits Alice wanted to conceal from Bob, are revealed after just a few milliseconds after she deposits them. Much faster than the  $10^{67}$  years she had hoped for. Because Bob holds the system  $E$  and knows the right decoding map to apply to a subsystem of  $\tilde{A}E$ , the black hole seems to behave like an information mirror and Alice qubits are bounced right back. Therefore, putting ourselves in an immediate scrambling situation due to the age of the black hole (it has already radiated away more than half of its initial entropy), we should ask: if Alice's quantum state persists behind the horizon and that state is also encoded in the outgoing Hawking radiation received by Bob, can Alice or Bob *verify* the cloning? In fact, we know from No-cloning theorem in 1.2.6, that no quantum xeroxing machines are allowed to exist, but the analyzed situation seems to violate this theorem. The answer to the question is complex, but we limit ourselves to assume that no violation of the No-cloning theorem occurs if no one is capable to demonstrate it. We have deeply talked about this issue in the section A.2 of the Appendix, where we introduced the *black hole complementarity* concept as a way to avoid this quandary. In fact, one chooses not to be bothered by quantum cloning if it occurs where no one can ever find out. According to this philosophy, we may accept for now that Alice (if she falls into the black hole) and Bob (if he stays outside) have sharply contrasting descriptions of the same physical process, both corresponding to the truth.

In the case Alice would have thrown her qubits into a young black hole (*i.e.*, before half of the black hole's initial entropy has been radiated away) we have that maximal entanglement between  $E$  and  $B$  does not occur yet and the initial state  $B$  is a pure state, then  $\sigma_{ABR_A}$  is also pure, and the Haar-averaged  $L^1$  distance of  $\sigma_{\widetilde{ABR_A}}$  from a product state is bounded above by

$$\sqrt{\frac{|ABR_A|}{|\tilde{A}|^2}} = \frac{2^{n+k}}{|2^{2s}|} = 2^{-c} \quad (3.17)$$

after  $s = \frac{1}{2}(n+k) + c$  qubits are emitted. This means that only after having waited for  $k+2c$  more qubits than the ones still residing in the black hole, he can succeed in decoding Alice's  $k$  qubits with fidelity  $F \geq 1 - 2^{-c}$ . In fact, Alice's  $k$ -qubit system has nothing special and Bob can decode any  $k$ -qubits he chooses from among the  $n$  of the initial black hole. That because the system  $R_A$  will however be very nearly maximally mixed with a  $k$ -qubit subsystem of  $\tilde{A}E$ . All Bob has to do is to chose  $k$  qubits of  $\tilde{A}E$  and perform a decoding operation on  $\tilde{A}E$  that maps those  $k$ -qubits to the system  $A$ .

Although there is far more to say about how black holes behave in quantum information processes that involves them, we will not delve further into this topic.

# Conclusions and Overviews

The purpose for this work is to offer some knowledge of fundamental aspects of quantum information theory. We tried to reach this by giving a brief introduction of quantum mechanical aspects that are crucial for understanding information theory. Later, we have analyzed how it is possible to pass from classical to quantum information theory, underlining how many of the central points of classical information theory are found to have quantum analogs. Starting from this, we have illustrated bounds on entropy, and therefore on classical information, encoded in quantum systems. Introducing quantum channels, we have also discussed bounds on quantum information sent reliably over a noisy quantum channel. In the last situation, we studied information retrieval from evaporating black holes, assuming that the internal dynamics of a black hole is unitary and rapidly mixing, and assuming that the retriever has unlimited control over the emitted Hawking radiation. What came out is that if the black hole has already evaporated more than half of its initial degrees of freedom, then additional quantum information deposited in the black hole is revealed in the Hawking radiation very rapidly. Information deposited prior to this point remains concealed until the previous situation occurs, and then emerges quickly.

We also have seen how black holes physics is strongly related to quantum information theory. In fact, we saw for Bekenstein entropy bound that black hole physics and thermodynamics allows us to reach more easily an important bound on information. But also in the last 'mirror black hole' application information theory allows us to use powerful tools and principles in order to reach conditions for black hole to be approximately considered a mirror for information, under certain conditions about the radiative process.

There is far more to say about quantum information theory and about the utility that tools developed for studying informative processes are used in quantum communication theory and other areas. For instance, all what we discussed about quantum channel capacity and about the decoupling principle provides a method to find achievable rates for certain quantum protocols or scheme for entangled-assisted quantum communication. Moreover, this introduces to quantum computation. We can be confident, though, that the concepts and applications discussed in this work leads to a good, though basic, comprehension of aspects of quantum information theory, hoping that it lights up the curiosity and passion for further studies.



# Appendix A

## Summary of black holes physics

In the previous section, we considered a highly idealized model of black hole dynamics using results usually motivated by gravitational physics, but formulated without reference to gravitation [15]. Therefore, this appendix wants to suggest another way to approach black holes entropy, some of the entropy bounds that have been inferred from it, and furnish a brief description of black hole thermodynamics. In addition to this, is discussed a universal relation between geometry and information; this also allows to mention different point of views assumed by scientific debate to interpret these arguments. The entropy bound discussed are independent of the specific characteristics and composition of matter systems. However, they apply only when gravity is weak. Throughout this appendix, Planck length will be used

$$l_p^2 = \frac{G\hbar}{c^3} = 2.59 \cdot 10^{-66} \text{cm}^2$$

(a thousand-trillion-trillion-trillion-trillion-trillion Plank areas would fit on the surface of a proton.) Here  $G$  is Newton's constant,  $\hbar$  is Planck's constant and  $c$  is the speed of light. Also Boltzman's constant  $k_B$  is used in the following lines.

### A.1 Black hole thermodynamics

Black hole is basically a region of space-time separated from the rest by a sort of one-way surface called *event horizon*. If you cross this horizon into this strongly gravitating region, there is no way out. Vice versa, if you stay outside this region, you will fine. Essentially, a place where information falls in and became inaccessible after passing through the horizon. The notion of black hole entropy is motivated by two results in general relativity [16].

**Theorem A.1.1** (Area). The area of a black-hole event horizon never decreases with time  $\delta A \geq 0$ .

The demonstration uses topological techniques beyond the scope of this work. Moreover, if two black holes merge, the final area will exceed the sum of the initial

areas. The theorem suggests an analogy between black hole area and thermodynamic entropy, as both are never decreasing quantities. Regarding this area increase as a kind of compensation for the loss of entropy of matter entropy threw into a black hole, Bekenstein suggested that a black hole actually carries an entropy equal to its horizon area, reaching the formula [17]

$$S_{BH} = \frac{A}{4} \quad (\text{A.1})$$

where the area is expressed in Planck units. We have already seen (section 2.2.2) that, according to Bekenstein, the entropy in a *bounded spatial region* must be  $S = O(ER)$  (with  $E$  the energy of matter enclosed in this region, and  $R$  the spherical radius of the region). Otherwise, one could violate the second law of thermodynamics by throwing extra material into the region and still obtain a final-state entropy level lower than the initial one. In the next section, we will see that this boundary also could be expressed as a *spherical entropy bound* for matter systems, thanks to (A.1).

**Theorem A.1.2** (No hair). A stationary black hole is characterized by only three quantities: mass  $M$ , angular momentum  $J$ , and charge  $Q$ .

From this point, is possible to generalize classical laws of thermodynamics in the context of black holes mechanics. This extension yield to the reformulation of the zeroth, the first and the third law, as the second law is already encoded in the Area theorem.

**Theorem A.1.3** (Zeroth Law or superficial gravity costance). A black hole has the same *superficial gravity*  $\kappa$  everywhere on the event horizon, it remains constant [18].

In fact, it depends only on  $M, J, Q$ . The definition of  $\kappa$  used in following lines is  $\kappa = (4M)^{-1}$ . Here the analogy with the Zeroth principle of thermodynamics is between  $\kappa$  and the the quantity called *temperature* which remains constant in a situation of thermodynamic equilibrium.

**Theorem A.1.4** (Third Law). To convert an ordinary black hole ( $\kappa \neq 0$  to an extremal one (for which  $\kappa = 0$ ), requires infinite steps [19].

This is in analogy to the Nernst law, which states that the temperature of absolute zero cannot be reached with finite number thermodynamic transformations. Further, as black holes have a mass  $M$  (no hair theorem) and entropy  $S$ , they also must have a *temperature*  $T$

$$dM = TdS_{BH} \quad (\text{A.2})$$

Einstein's relations imply

**Theorem A.1.5** (First Law).

$$dM = \frac{\kappa}{8\pi} dA \quad (\text{A.3})$$

where the entropy is the horizon area and  $\kappa$  plays the role of temperature.

At first, was this similarity between (A.2) and (A.3) to suggest an identification between quantities characterizing black holes and thermodynamic quantities.

$$T \propto \kappa, \quad S \propto A$$

However, no physical meaning has long been given to the proportions above, since a black hole equipped with  $\kappa$ , also must have a non-zero *temperature*, so it must radiate. Conversely, a black hole is by definition black body, which absorbs every radiation thrown into it without radiating, so its *temperature* must be exactly zero. In addition to this, standing Nernst theorem, a body at absolute zero has null entropy  $S_{BH} = 0$ . So, a process of a complex matter system collapsing to form a final-black-hole state appears to violate the second law of thermodynamics. Hence, Bekenstein [3] proposed that the second law of thermodynamics holds only for the sum of matter entropy  $S_{ext}$  and black-hole entropy  $S_{BH}$

**Theorem A.1.6** (Generalized Second Law).  $\delta(S_{ext} + S_{BH}) \geq 0$

The *GSL* still remains a formal argument, for reasoning above. In fact, according to thermodynamics, a black hole with entropy  $S_{BH}$  should radiate with a black-body spectrum corresponding to the temperature

$$T_{BH} = \alpha \cdot \kappa \tag{A.4}$$

Throughout the dimensional analysis it turns out that the constant  $\alpha$  should have dimensions of a temperature multiplied by a mass [16]. By solely invoking  $k_B$ ,  $G$ , and  $c$  in calculations, is impossible to fulfil this requirement. Similarly, to satisfy the proportionality between entropy and area, a constant  $\beta$  is needed.

$$S_{BH} = \beta \cdot A \tag{A.5}$$

As entropy has the dimensions of  $k_B$ , this constant must be an energy dividing a temperature and a length squared. Once again, only  $k_B$ ,  $G$ , and  $c$  do not suffice. So, dimension of  $\hbar$  is needed. In fact, looking for quantum corrections to black-hole physics, Hawking found that black holes do really radiate with a spectrum at a temperature given by (A.4) [17]. So, the entropy and temperature of a black hole are no less real than its mass. Another process related to this is the Hawking evaporation of a black hole, explained in the next section. Although these types of processes were not anticipated when Bekenstein proposed black hole entropy and the *GSL*, it is calculated that *GSL* holds also in these cases [3].

### A.1.1 Bekenstein and spherical bound for entropy of matter systems

Here, we will see a different way to address the problem of limiting entropy and information into an enclosed space, with respect to the one in section 2.2.2. When information is dropped into a black hole, its entropy  $S_{ext}$  vanishes to an external

observer. But the entropy of a black hole increases in virtue of (A.1), because the black hole gains mass, then area. Thus, it is at least conceivable that total entropy  $\delta(S_{ext} + S_{BH})$  does not decrease in the process. Though, to ensure that GSL also holds in this situation, should exist a universal bound on the entropy of the matter system. Consider a weakly gravitating matter system of total energy  $E$ . Let  $R$  be the radius of the smallest sphere that contains the system. We want to collapse our system into a black hole adding as little energy as possible to it, so as to minimize the increase of the horizon's area and thus optimizing the tightness of the entropy bound (*Geroch process*). Using this idea, it turns out that black-hole entropy increases by

$$\delta S_{BH} \leq \frac{2\pi k_B ER}{\hbar c}$$

By the GSL, this increase must at least compensate for the lost matter entropy:  $\delta S_{BH} - S_{ext} \geq 0$ . Hence, Bekenstein argued [20]:

$$S_{ext} \leq \frac{2\pi k_B ER}{\hbar c} \tag{A.6}$$

That is the ***Bekenstein bound***, in a similar formulation to the one given in Chapter 3. Note that Newton's constant does not enter. Instead of dropping a thermodynamic system into an existing black hole via the *Geroch process*, one may also consider the *Susskind process*, in which the system is converted to a black hole. In this case we consider a system of a certain mass enclosed in a circumscribing sphere of area  $A$ . The mass of the system must be less than the mass  $M$  of a black hole with the same horizon area  $A$ . Otherwise, the system would also be collapsed into a black hole. We expect the system converts into a black hole of surface  $A$  by collapsing a shell of mass  $M - E$  onto the system. The total initial entropy is only given by  $S_{ext}$  of the system, while the final one is a black hole, given by (A.1). As the initial entropy must not exceed the initial entropy, holds that:

$$S_{ext} \leq \frac{A}{4} \tag{A.7}$$

That is the ***spherical entropy bound***, which is saturated only by a black hole, so it is the most entropic object one can put inside a given spherical surface.

In order to be more accurate, we could speak about information rather than systems made of matter. It is important to see that exists an upper limit for the information stored in an enclosed space. Consider an empty room (except air), one can associate pieces of information to air molecules inside this space. To increase the amount of information, one could tightly pack up these molecules against each other so they cannot be treated as a gas anymore. Going further, molecules can be broken up into atoms and then nuclei and even into smaller things such as protons and neutrons and quarks. In quantum field theory there is no limit to this process, and one could reach every infinitesimal scale; so, there is an infinite amount of information in any region of space. The uncertainty principle  $\Delta E \Delta t \sim \hbar$



explains that the amount of energy needed to build a feature of size  $\Delta x$  is inversely proportional to that size, so in our analogy smaller letters are heavier. Finally, we reach the general result that the more information we want to put into our room, the more energy we have to put into the room (and is found that the energy increases dramatically as I refine my resolution). Up to this point we have neglected gravity, which really poses a limit on the amount of mass you can have in the room, just by specify how large the room is in light of spherical entropy bound. In fact, increasing more and more the mass in a given volume we face gravitational instability, then a *Susskind process* occurs and the room is converted into a black hole. So, we obtained there is an upper limit to the energy in the room, then a limit on the entropy, consequently a limit on the amount of information that we could store in the room. There are general statements that tell us about the information at most expected to store in a region, irrespective of the system considered. In fact, black holes have an entropy, then an information content, and it is given by the surface area of the horizon. So, in a general context, every boundaries discussed above are determined by (A.1). We will explore more in detail this point in the subsection A.1.3.

### A.1.2 Hawking radiation and evaporation

The solution to the entropy paradox was given by the discovery that black holes radiate via quantum process [17]. Hawking showed by semi-classical calculation of quantum theory in curved spaces, that the temperature corresponding to the emitted black-body spectrum is of the same type of (A.4). The point is that, according to quantum field theory, the void is characterized by quantum fluctuations which originate virtual couples of particle-antiparticle nonstop created and immediately destroyed. This virtual processes holds by virtue of Heisenberg uncertainty principle, which allows to force-mediating particles to have high quantities of energy for little time. Assuming that a couple is created in close proximity to event horizon. One particle has positive energy  $E$ , while the other has negative energy  $-E$ . Now, we consider that generally the couple is annihilated within a time interval  $\Delta t$ , but next to the event horizon could happen that the particle with negative energy crosses the horizon earlier than  $\Delta t \sim \frac{\hbar}{E}$  pass. Beyond the horizon the negative energy particle will reduce the black-hole total energy  $M$  by a value  $E$ . Whereas, the positive energy particle can freely escape on the outside, constituting a detectable radiation. For a black hole, the Schwarzschild gravitational potential represents a barrier to overcome by particles in order to escape from black hole. In this case a virtual couple is converted into a real one at the expense of black-hole energy. This could happen by tunnel effect, a process exponentially dependent on the height of the barrier. So particles with  $l \neq 0$  have very few probability to run away. Therefore, radiation will result in more than 90% of  $l = 0$  particles [16].

Surprisingly, it turned out that these probabilities of particles succeeding in going

beyond the horizon by tunnel effect, furnish a distribution

$$N(\omega) = \frac{1}{e^{8\pi\omega GM/c^3} - 1} \quad (\text{A.8})$$

of the form of the Planck distribution for a boson thermal radiation.

$$N(\omega) = \frac{1}{e^{\hbar\omega/k_B T_{BH}} - 1}$$

Therefore, the black body must radiate at a temperature

$$T_{BH} = \frac{\hbar c^3}{8\pi GM k_B} \quad (\text{A.9})$$

Finally, (A.4) and (A.5) can be reformulated with constants exploited

$$T_{BH} = \frac{\hbar c^3 \kappa}{2\pi G k_B} \quad (\text{A.10a})$$

$$S_{BH} = \frac{k_B c^3 A}{G \hbar 4} \quad (\text{A.10b})$$

So, in units of Plank length squared, from the second (A.10b) we recover (A.1), demonstrating again that black holes really radiate and this process yields to Bekenstein entropy bound. A general rule is that the more a black hole is massive, the less it irradiates. Therefore, its temperature is intended to increase and its mass to decrease, boosting the phenomenon of radiation. We can estimate the lifetime of a black hole by imaging the whole process as quasi-static. According to Stefan-Boltzmann law, the radiant power is proportional to the radiant Area  $A$  and to  $T_{BH}^4$ . Considering also that a particle steal an energy of  $\hbar/8\pi M$  from the black hole, we reach a differential equation describing the variation of the black hole's mass with respect to time

$$\frac{dM}{dt} = -\gamma \frac{m_p^3}{t_p} \frac{1}{M^2} \quad (\text{A.11})$$

where  $m_p = \sqrt{\hbar c/G} \sim 10^{19} GeV$  is the Planck mass,  $t_p = \sqrt{\hbar G/c} \sim 10^{-44} sec$  is the Planck time and  $\gamma \sim 10^{-5}$  is an adimensional constant. Integrating (A.11), we obtain

$$M(t) = \left( M_0^3 - \frac{3\gamma m_p^3}{t_p} t \right)^{1/3} \quad (\text{A.12})$$

and therefore the black hole will be evaporated in a time

$$t_{ev} = \frac{t_p}{3\gamma} \left( \frac{M_0}{m_p} \right)^3 \quad (\text{A.13})$$

So, the evaporation time is  $t_{ev} \sim 10^{66-69}$  years for stellar black holes with typical masses of 3 – 10 times solar mass. Consequently, one could say that the time involved by a black hole in expelling a significant fraction of qubits (*i.e.*, radiate away a significant part of its mass) is of the order  $O(n^{3/2})$ .

### A.1.3 Quantum states of a black hole and thermalization

The number of degrees of freedom of a quantum-mechanical system  $N$  is the logarithm of the dimension  $\mathcal{D}$  of its Hilbert space  $\mathcal{H}$ :

$$N = \ln \mathcal{D} = \ln \dim(\mathcal{H}) \quad (\text{A.14})$$

Entropy measures the number of allowed quantum states for a system. For a black hole that just fits inside a region with area  $A$ , its entropy is given by (A.1) and this clearly saturates the spherical entropy bound (A.7) for a system, but more generally for an enclosed spacial region. So, the number of degrees of freedom in a region bounded by a sphere of area  $A$  is given by

$$N = \frac{A}{4}$$

Therefore, the number of states is (*Bekenstein limit*)

$$\mathcal{D} = e^{A/4} \quad (\text{A.15})$$

Hence, using the spherical entropy bound, we have concluded that  $A/4$  degrees of freedom are sufficient to fully describe any system enclosed by a sphere  $A$ . It is also possible to demonstrate that any attempt to excite more than  $A/4$  of these degrees of freedom is thwarted by gravitational collapse. From an external point of view, the most entropic object that fits in an enclosed space is a black hole of area  $A$ , with  $A/4$  degree of freedom.

All this information seems to be distributed in a random manner over the horizon surface. In fact, considering that black holes do have a temperature, behind the black body distribution of the frequency of the emitted particles by Hawking radiation, there might conceal a thermal agitation of this degrees of freedom which causes for them to be in a constant state of agitated motion, to be very chaotic and so causing for the information to be randomly distributed among the horizon surface. From a certain point of view, the information inside a black hole could run into a 'thermalization' process that is worthy to be studied in depth. It consists of simultaneous destruction and random recombination in the form of uncorrelated degrees of freedom in the event horizon surface of an information signal, that later will be radiated away in the form of black body radiation. Is important to consider that this processes can happen so rapidly that they can be considered as one strongly related to thermal agitation and black hole's temperature, hence we talk about thermalization. Moreover, the sum of all this degrees of freedom conveys the same information as before running into black hole, the difference is that now everything has been rapidly mixed up by the internal dynamics of the black hole. This result is known as the *fast scrambling conjecture*. Moreover, relying on this conjecture, for a black hole of temperature  $T$ , each qubit takes a time of order of  $\hbar/kT$  to be emitted and a black holes takes a time logarithmic in its entropy  $O(\ln[\dim(\mathcal{H})])$  to scramble information, and no system in nature can scramble faster [21]. Therefore, from this

point of view, it seems that information inside a black hole is randomly distributed among the horizon surface in the form of a *disordered hologram* of what originally was tossed into the black hole. From this point take place the holographic principle, but to give a complete and accurate description of this is far beyond the aim of this appendix, therefore we limit ourselves to say that in a certain way, the black hole horizon might be considered as an hologram of what is inside.

## A.2 The problem of unitarity and black hole complementarity

It seems fascinating to investigate what really happens when final stages of black hole evaporation occurs. In fact, at some point the mass  $M$  become of the same magnitude of the Planck mass  $m_p$  with a size of Plank length cubed. At this point a great conceptual paradox rises. The matter system, which collapses to form a black hole, contains a great amount of information. This seems to vanish once crossed the horizon. One could argue that the Hawking radiation process returns this information back, but it is not so. This process constitutes of thermal isotropic-scattering particles, so does not contain ordered information. Thus, it appears that information is lost when threw into a black hole and this cannot happens, because of the principle of conservation of information. A more compelling consideration deals with *quantum unitarity*. Quantum-mechanical evolution preserves information, as it takes a pure state to a pure state. But, consider a region described by a Hilbert space of dimension  $e^V$ , and suppose that region was converted to a black hole. According to Bekenstein limit, the region is now described by a Hilbert space of dimension  $e^{A/4}$  (with  $V > A/4$ ). The number of states would have decreased, and it would be impossible to recover the initial state from the final one. Thus, it seems that unitarity is not preserved in the presence of black holes. However, unitarity must be restored in a complete quantum gravity theory. There have been many answers to this question and part of the debate developed around this is well known as '*the black hole war*'. From one side we have Hawking, who has claimed for years that a quantum theory of gravitation could violate unitarity. In fact, is not understood in detail how Hawking radiation carries away information. However, it seems inevitable that the evaporation of a black hole-its slow conversion into a cloud of thermalized radiation is not a unitary process. From the other side, Susskind, Thorlacius and Uglum (*et al.* such as Preskill and Page) resolved the information paradox arguing that *when a black hole evaporates unitarily, the same quantum information would seem to be present both inside (as the original system matter that collapsed) and outside the black hole (in the form of Hawking radiation)* [15]. However, the simultaneous presence of two copies could not be possible because of the No Cloning theorem, which forbids the 'xeroxing' of information. The upshot is that there are two different complementary descriptions of black-hole, corresponding to an infalling and an outstanding observer. It turns out that each point of view is self-consistent, since

no single observer can see both copies of the information simultaneously. But, a description of these views is neither logically consistent nor practically testable.

To understand, we could consider two friends Alice and Bob sharing two copies of the same code. We assume Alice could cross the horizon and continue to fall recognizing nothing out of the ordinary; but also, Alice could be thermalized (via Hawking radiation process) at the horizon and then radiated back out as photons visible to Bob, the outstanding observer. As stated, both these realities are true. Let's push this situation a little bit further by assuming that the thermalization occurs just as Alice crosses the horizon. In this moment Bob takes a look at Alice in order to see her copy of the code. This means that he shines electromagnetic radiation on her which bounces off into his eyes. Because of gravity increase as horizon approaches, the more photons are shipped near the horizon, the more it takes them to return back and their frequency decrease. That because of gravity. Exactly in the horizon their redshift is infinite: all their energy is used to escape from gravitational field. So, Bob has to shot Alice with short enough wavelengths photons (to resist redshift and return to him) that will themselves *thermalize* her and her message. Furthermore, assume that Alice quietly crosses horizon and Bob wants to jump into the black hole to read Alice's copy of the code. The second copy can only be observed if it has not already hit the singularity inside the black hole by the time Bob crosses the horizon. It is showed that the energy required for a single photon to evade the singularity for so long is exponential in the square of the black hole mass. In other words, there is far too little energy in the black hole to communicate even one bit of information to an infalling observer in possession of outside data. Note that an important role is now given to observer also in quantum gravity. The *black-hole complementarity* is thus exposed and explained as a way to interpret the formation and evaporation of a black hole as a unitary process, at the expense of locality.

Anyway, is studied that black-hole complementarity holds in asymptotically flat spacetimes. So, seems that a complete answer to information paradox has already to be given. Recent developments regards the framework of quantum states entanglement as the one from which answers to this problem could be found.



# Bibliography

- [1] C. E. Shannon, “A mathematical theory of communication,” *Bell System Technical Journal*, July/October 1948. <https://web.archive.org/web/19980715013250/http://cm.bell-labs.com/cm/ms/what/shannonday/shannon1948.pdf>.
- [2] W. Wootters and W. Zurek, “A single quantum cannot be cloned,” *Nature*, vol. 299 (5886): 802–803, 1982. doi:10.1038/299802a0. S2CID 4339227.
- [3] J. Bekenstein, “Black holes and the second law,” *Let. Nuovo Cimento Soc. Ital. Fis.*, vol. 4:737-740, May 1972. [http://old.phys.huji.ac.il/~barak\\_kol/Courses/Black-holes/reading-papers/Beken-Entropy.pdf](http://old.phys.huji.ac.il/~barak_kol/Courses/Black-holes/reading-papers/Beken-Entropy.pdf).
- [4] John Preskill, “Lecture Notes for Ph219/CS219: *Chapter 1. Introduction and Overview.*” <http://theory.caltech.edu/~preskill/ph229/notes/chap1.pdf>, 1997. California Institute of Technology.
- [5] John Preskill, “Lecture Notes for Ph219/CS219: *Chapter 2. Foundation I: States and Ensembles.*” [http://theory.caltech.edu/~preskill/ph219/chap2\\_15.pdf](http://theory.caltech.edu/~preskill/ph219/chap2_15.pdf), July 2015. California Institute of Technology.
- [6] John Preskill, “Lecture Notes for Ph219/CS219: *Chapter 3. Foundations II: Measurement and Evolution.*” [http://theory.caltech.edu/~preskill/ph219/chap3\\_15.pdf](http://theory.caltech.edu/~preskill/ph219/chap3_15.pdf), July 2015. California Institute of Technology.
- [7] John Preskill, “Lecture Notes for Ph219/CS219: *Chapter 4. Quantum Entanglement.*” [http://theory.caltech.edu/~preskill/ph229/notes/chap4\\_01.pdf](http://theory.caltech.edu/~preskill/ph229/notes/chap4_01.pdf), 2001. California Institute of Technology.
- [8] John Preskill, “Lecture Notes for Ph219/CS219: *Chapter 10. Quantum Shannon Theory.*” [http://theory.caltech.edu/~preskill/ph219/chap10\\_6A.pdf](http://theory.caltech.edu/~preskill/ph219/chap10_6A.pdf), January 2018. California Institute of Technology.
- [9] P. Hayden and J. Preskill, “Black holes as mirrors: quantum information in random subsystems,” *CalTech press*, September 2007. <https://arxiv.org/pdf/0708.4025.pdf>.
- [10] S. Hawking, “Breakdown of predictability in gravitational collapse,” *Physical review*, vol. 14, 2460, November 1976.

- [11] P. Dirac, “The principles of quantum mechanics,” *Clarendon Press*, vol. 2nd ed., 1947.
- [12] R. Yuan, “A brief introduction to povm measurement in quantum communications,” *Beijing University Press*. <https://arxiv.org/ftp/arxiv/papers/2201/2201.07968.pdf>.
- [13] P. Algoet and T. Cover, “A sandwich proof of the shannon-mcmillan-breiman theorem,” *The Annals of Probability*., vol. 16 (2): 899–909, 1988. <https://isl.stanford.edu/~cover/papers/paper83.pdf>.
- [14] P. Hayden, M. Horodecki, A. Winter, and J. Yard, “A decoupling approach to the quantum capacity,” *Open Systems Information Dynamics*, vol. 15(1):7-19, March 2008a. <https://arxiv.org/pdf/quant-ph/0702005.pdf>.
- [15] R. Bousso, “The holographic principle,” *Review of Modern Physics*, vol. 74:825-874, 2002. <https://arxiv.org/pdf/hep-th/0203101.pdf>.
- [16] Francesco Ravanini, “Appunti di relatività,” March 17, 2014. Lecture notes.
- [17] S. Hawking, “Black hole explosions,” *Nature*, vol. 248, pp. 30–31, March 1974.
- [18] J. Bardeen, B. Carter, and S. Hawking, “The four laws of black hole mechanics,” *Comm. Math. Phys.*, vol. 31(2): 161-170, 1973. <https://projecteuclid.org/journals/communications-in-mathematical-physics/volume-31/issue-2/The-four-laws-of-black-hole-mechanics/cmp/1103858973.full>.
- [19] W. Israel, “Third law of black-hole dynamics: A formulation and proof,” *Physical Review*, vol. 57:397, July 1986. <https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.57.397>.
- [20] J. Bekenstein, “Universal upper bound on the entropy-to-energy ratio for bounded systems,” *Physical Review*, vol. 23:287, January 1981. <https://journals.aps.org/prd/abstract/10.1103/PhysRevD.23.287>.
- [21] N. Lashkari, D. Stanford, M. Hastings, T. Osborne, and P. Hayden, “Towards the fast scrambling conjecture,” *Cornell University press*, November 2011. <https://arxiv.org/pdf/1111.6580.pdf>.
- [22] John Preskill, “Lecture Notes for Ph219/CS219: Chapter 5. Quantum Information.” [http://theory.caltech.edu/~preskill/ph219/chap5\\_15.pdf](http://theory.caltech.edu/~preskill/ph219/chap5_15.pdf), July 2015. California Institute of Technology.
- [23] W. Heisenberg, “Über den anschaulichen inhalt der quantentheoretischen kinematik und mechanik,” *Zeitschrift für Physik*, vol. 43 (3–4), 1927.



- 
- [24] M. Macon, “The Noisy Channel Coding Theorem.” <http://math.sfsu.edu/serkan/expository/michaelMaconExpository.pdf>, December 18, 2015. Expository.