CBPF 70 ANOS

MINISTÉRIO DA CIÊNCIA, TECNOLOGIA, INOVAÇÕES E COMUNICAÇÕES

A review on Quantum information and Black Holes

a dissertation presented

by

João G. A. Caribé

to

The Department of Astrophysics, Cosmology and Fundamental Interactions

in partial fulfillment of the requirements
for the degree of
Master of Physics
in the subject of
Physics

Centro Brasileiro de Pesquisas Físicas
Rio de Janeiro, Rio de Janeiro
August 2019

CBPF 70 ANOS

MINISTÉRIO DA CIÊNCIA, TECNOLOGIA, INOVAÇÕES E COMUNICAÇÕES

"A REVIEW ON QUANTUM INFORMATION AND BLACK HOLES"

JOÃO GABRIEL ALENCAR CARIBÉ

Dissertação de Mestrado em Física, apresentada no Centro Brasileiro de Pesquisas Físicas do Ministério da Ciência, Tecnologia, Inovações e Comunicações. Fazendo parte da banca examinadora os seguintes professores:

Marc Casals Casanellas- Presidente/Orientador/CBPF

Sérgio Eduardo de Carvalho Jorás - UFRJ

Fernando da Rocha Vaz Bandeira de Melo - CBPF

Rio de Janeiro, 05 de novembro de 2019.

Acknowledgments

I would like to thank my family for all the support during this project. Without them this path would have been much tougher. I thank my friends, that helped me during this project with both, good discussions and moments of happiness. I also thank Carol for being comprehensive during all this time.

I am also grateful by being advised by Marc. These two years were very rewarding and I learned a lot about both, Physics and being a researcher. In every single meeting we had, I always left with more questions which stimulated me to learn more. I also remark that Marc had a HUGE amount of patience while teaching me a specific point of the English grammar. Thanks Marc.

I also thank Helayël, who played a fundamental role during my very first day at CBPF.

I thank all staff and professors for making CBPF such a great place to be.

Last, but not least, I thank every single professor I had in my whole life. Without their teachings, this would have not been possible.

During this Masters project, I was beneficiary of a CAPES(Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) studentship

Thesis advisor: Professor Marc Casals Casanellas João G. A. Caribé

A review on Quantum information and Black Holes

Abstract

The quantum dynamics of a black hole is still an open problem. Currently, the framework of semi-classical General Relativity is the most popular approach towards implementing quantum effects on spacetime. While being theoretically consistent and well accepted, it leads to a paradox when one tries to apply it to black holes. By using the semi-classical framework, one gets to the conclusion that black holes evaporate into thermal radiation, apparently erasing any information about matter that fell into it. That problem is particularly interesting because it links many areas of physics, namely, Quantum Information, Statistical Mechanics and General Relativity, which makes it fascinating but also hard to understand. In this work we review the minimal necessary formalism needed to state the paradox and explain two statements for it. In the last chapter we also present a toy-model that behaves in a way that resembles what is expected from an evaporating black hole. We believe that by further developing such model it might be possible to model more precisely what is expected from an evaporating black hole.

Keywords: Hawking radiation, Information, Paradox

A review on Quantum information and Black Holes

Resumo

A dinâmica quântica de um buraco negro ainda é um problema em aberto. Atualmente, o formalismo da gravitação semi-clássica é o método mais popular de incluir efeitos quânticos no espaço-tempo. Mesmo sendo teoricamente consistente e bem aceito, esse formalismo leva a um paradox quando apicado a buracos negros. Usando o formalismo semi-clássico, somos levados a conclusão de que buracos negros evaporam por meio de radição térmica, aparentemente apagando informação sobre tudo o que foi caiu dentro dele. Esse problema é particularmente interessante porque atravessa diversas áreas da Física, especificamente, Mecânica Quântica, Mecânica Estatística e Relatividade Geral, o que o torna fascinante mas ao mesmo tempo difícil de entender. Nesse trabalho fazemos uma revisão do formalismo mínimo necessário para enunciar o paradox e também explicamos dois enunciados para o mesmo. No último capítulo, apresentamos um *toy-model* que se comporta de maneira parecida com o que se espera de um buraco negro evaporando. Acreditamos que desenvolvendo esse modelo, talvez seja possível modelar de maneira mais precisa o que se espera de um buraco negro evaporando.

Palavras-chave: Radiação Hawking, Informação, Paradoxo

Contents

Listing of figures

0

Introduction

In the early '70s, Hawking showed that the area of a Black hole event horizon cannot decrease when undergoing classical physical processes[1]. Two years after Bekenstein derived a relation between a Black Hole mass, energy, area and charge that resembled the second law of thermodynamics and also proposed a thermodynamic entropy for Black Holes, which in the same work was estimated to be proportional to Black Hole area[2]. Such a result is credible because the area of a Black hole presents the same behavior as the thermodynamic entropy of ordinary physical systems: Both cannot decrease when undergoing classical physical processes.

Bekenstein's work was at first sight questioned by Hawking because according to thermodynamics, if one can define a thermodynamic entropy for a physical system, then it must also be possible to define a temperature. At the same time, anything that has a non-zero temperature must be radiating heat. But up to that time there wasn't any known mechanism allowing Black Holes to radiate. However, such an argument only held up for two years.

In 1974 Hawking considered a quantum massless scalar field on a collapsing star background and found out that in this setup radiation is expected to come out from the Black Hole obeying a spectrum that is the same as it would be if the Black Hole were an ordinary hot body[3]. From this time on, Bekenstein's ideas on Black Hole entropy gained substance on theoretical grounds, as now at least one mechanism is known to allow for Black Holes to radiate which in turn solves the questioning present on the previous paragraph.

Two years later Hawking presented a new article where he argued that there is a breakdown of predictability in gravitational collapse[4] in the sense that there is no S matrix describing the process of Black Hole evaporation, therefore, information about the state of the matter before the collapse is lost. In the same article, he proposed that there would in principle be no problem with the information loss and proposed a non-unitary operator called the superscattering operator, which would map the initial to the final state of the system.

Banks, Susskind & Peskin gave a strong argument against the superscattering approach

[5]. They showed that such an approach would violate either locality by creating correlations between widely separated points on the same space-like hypersurface or energy-momentum conservation none of which is desirable because both would have catastrophic physical consequences like faster than light communication and unbounded violation of energy conservation.

In 1987 another proposal for a solution to the information loss problem came in an article by Aharonov, Casher & Nussinov[6]. They proposed for the first time that a Black hole might stop evaporating after reaching Planck scale, leaving a long-lived remnant. The main difficulty for remnants is that they should be ubiquitous objects due to their high degeneracy and very low energy, but they have not yet been observed.

The remnant proposal was in the center of the discussion until 1993, when Susskind, Thorlacius & Uglum proposed the idea of Black hole complementarity [7]. That solution consists of a set of 4 postulates concerning the unitarity of quantum mechanics, the equivalence principle of General Relativity, Bekenstein's proposal for Black hole entropy, and the membrane paradigm. The latter is a formalism introduced by Price, Thorn & MacDonald [8], which describes a Black hole in terms of its surface properties. According to the complementarity hypothesis, observations made by someone inside a Black hole are dual to that made by someone outside a Black hole. That duality will be explored in section 4.4. The problem with complementarity is that it seems not to be enough and its postulates might be inconsistent as argued in [9].

Up to this day, those proposals were the ones that attracted most of the attention. The objective of this text is to present, being as much self-consistent as possible, a precise statement of the information loss puzzle and a brief explanation for those major proposals of solution. In the first Chapter 1, we present a rapid introduction to some concepts of Quantum Mechanics that are important for understanding the information puzzle.

In Chapter 2 we present to the reader some aspects of thermodynamics and statistical mechanics as they form the basis for the somewhat confusing concept of entropy, other than that some remarks about on subtleties of the laws of thermodynamics are exposed.

In Chapter 3, we present the basis of General Relativity and Black Hole Thermodynamics. The first is necessary to understand the foundations of the arguments for the information loss puzzle while the second originated the discussion. In the end, we briefly show, in a schematic way, how the semi-classical framework works.

Chapter 4 is the most important chapter. It contains the statement of the information loss puzzle and the explanations about some solution proposals. We finish in Chapter 5 with a motivation and a toy model aiming to reproduce, to some degree, the expected behavior of entanglement and thermodynamics during Black hole evaporation.

1

Quantum Mechanics

Being the basis for several arguments that will be presented throughout the thesis, we should begin with a quick exposition on Quantum Mechanics.

1.1 Postulates of quantum mechanics

To construct the framework of quantum mechanics, one can follow a set of postulates. The ones presented below result from combining the postulates presented in [10] with the more mathematical approach of [11].

Postulate 1.1.1. (Adapted from [10]) Associated to any isolated physical system there is a Hilbert space $\mathcal{H}$. The state space of the system is defined as the convex set

$$\mathcal{M} = \left\{\rho \mid \rho : \mathcal{H} \mapsto \mathcal{H}, \operatorname{Tr}\rho = 1, \rho^\dagger = \rho\right\}. \tag{1.1}$$

The elements of the state space are called density operators. A system is completely described by its density operator. If a quantum system is in a state $\rho_i \in \mathcal{M}$ with probability $p_i \in [0, 1]$, then the density operator for the system is

$$\rho = \sum_i p_i \rho_i. \tag{1.2}$$

Postulate 1.1.2. (Adapted from [10]) The evolution of a closed quantum system is described by a unitary transformation. That is, the state of the system at time $t_1 \in \mathbb{R}_{\geq 0}$ is related to the state at time $t_2 \in \mathbb{R}_{\geq 0}$ by a unitary operator $U : \mathcal{H} \mapsto \mathcal{H}$ which depends only on the times t_1 and t_2,

$$\rho(t_2) = U(t_1, t_2)\, \rho(t_1)\, U^\dagger(t_1, t_2). \tag{1.3}$$

Postulate 1.1.3. (As is from Nielsen&Chaung[10]) Quantum measurements are described by a collection $\{M_m\}$ of measurement operators $M_m : \mathcal{H} \mapsto \mathcal{H}$. These are operators acting on the state space of the system being measured. The index m refers to the measurement outcomes that may occur in the experiment. If the state of the quantum system is ρ immediately before the measurement then the probability that result m occurs is given by

$$p\left(m\right) = \mathrm{Tr}\left(M_m^\dagger M_m \rho\right), \tag{1.4}$$

and the state of the system after the measurement is

$$\frac{M_m^\dagger \rho M_m}{\mathrm{Tr}\left(M_m^\dagger M_m \rho\right)}. \tag{1.5}$$

The measurement operators satisfy the completeness equation,

$$\sum_m M_m^\dagger M_m = \mathbf{1}. \tag{1.6}$$

Postulate 1.1.4. (Adapted from [10]) The state space of a composite physical system is the tensor product of the state spaces of the component physical systems. Moreover, if we have systems numbered 1 through n, and the system number i is prepared in the state $\rho_i \in \mathcal{M}_i$, then the joint state of the total system is $\rho_1 \otimes \rho_2 \otimes \ldots \rho_n \in \bigotimes_{i=1}^{n} \mathcal{M}_i$.

The state-space $\mathcal{M}$ is a subset of the set of operators acting on $\mathcal{H}$, the Hilbert-Schmidt space $\mathcal{HS}$, equipped with the inner product

$$\begin{aligned} \langle \cdot, \cdot \rangle &: \mathcal{M} \times \mathcal{M} \mapsto \mathbb{C}, \\ \langle A, B \rangle &= \mathrm{Tr}\left(A^\dagger B\right), \end{aligned} \tag{1.7}$$

and consequently, the inner product norm

$$\begin{aligned} \|\cdot\| &: \mathcal{M} \mapsto \mathbb{R}, \\ \|\rho\| &= \sqrt{\langle \rho, \rho \rangle} = \sqrt{\mathrm{Tr}\left(\rho^2\right)}. \end{aligned} \tag{1.8}$$

The norm is bounded on $\left[\dim\left(\mathcal{H}\right)^{-1}, 1\right]$ and the eigenvalues of density operators are bounded on $[0, 1]$. A proof for those claims can be found in appendixes A.2 and A.1. One then define pure and mixed states as follows:

Definition 1.2.1 (Pure state). Any physical system whose associated density operator $\rho \in \mathcal{M}$ is such that $\|\rho\| = 1$ is said to be in a pure state. Additionally, let $\mathcal{P} = \{\rho \mid \|\rho\| = 1\} \subset \mathcal{M}$ be the set of pure states contained in the state space of the system.

Definition 1.2.2 (Mixed states). Any physical system whose associated density operator $\rho \in \mathcal{M}$ is such that $\|\rho\| < 1$ is said to be in a mixed state.

There is only one state that minimizes the norm of a given state space, it is $\rho = \dfrac{\mathbf{1}}{\dim\left(\mathcal{H}\right)}$. Such a state is called the maximally mixed state.

There is a quantity that measures the mixedness of a quantum system; it is called purity and is defined as follows:

Definition 1.2.3 (Purity). The purity of a quantum state $\rho \in \mathcal{M}$ is defined as

$$\begin{aligned} \gamma &: \mathcal{M} \mapsto \mathbb{R} \\ \gamma(\rho) &= \operatorname{Tr}\left(\rho^2\right). \end{aligned} \tag{1.9}$$

The purity is clearly bounded in $\left[\dim(\mathcal{H})^{-2}, 1\right]$. The bigger the purity, the closer a quantum system is to being pure. Therefore, it is possible to order quantum systems with respect to their purity.

1.3 The Hilbert Space formalism

We began by presenting the density operator formalism of Quantum mechanics. Although not explicit, the usual Hilbert space formalism can be recovered, and this is the goal of this section.

To recover the Hilbert space formalism one should start by noting that $\dim(\mathcal{P}) = \dim(\mathcal{H})$ because there are $\dim(\mathcal{H})$ eigenvalues for any given pure state density . Hence, $\mathcal{P}$ and $\mathcal{H}$ are isomorphic because both have the same dimension and are vector spaces over the same field [12]. Because of that, it is possible to identify each pure state density operator with a unique vector on $\mathcal{H}$. Because of that, one can also refer to the elements of $\mathcal{H}$ as pure quantum states.

In order to do that, we first use the inner product structure to define functionals over $\mathcal{H}$

as

$$\langle\psi| : \mathcal{H} \mapsto \mathcal{C}$$
$$\langle\psi| \left(|\phi\rangle\right) := \left(|\psi\rangle , |\phi\rangle\right) \equiv \langle\psi|\phi\rangle , |\psi\rangle , |\phi\rangle \in \mathcal{H}. \tag{1.10}$$

For every $|\psi\rangle \in \mathcal{H}$ there is a unique $\langle\psi|$ as guaranteed by Riesz' Representation Theorem[13]. Moreover, the space of such functions over $\mathcal{H}$ is denoted $\mathcal{H}^*$. We now define the following map:

$$f : \mathcal{H} \mapsto \mathcal{H} \otimes \mathcal{H}^*,$$
$$f\left(|\phi\rangle\right) = |\phi\rangle \otimes \langle\phi| \equiv |\phi\rangle\langle\phi| , |\phi\rangle \in \mathcal{H}. \tag{1.11}$$

The action of an element of $\mathcal{H} \otimes \mathcal{H}^*$ on an element of $\mathcal{H}^*$ is defined as

$$|\phi\rangle\langle\phi| \left(|\psi\rangle\right) := \langle\phi|\psi\rangle |\phi\rangle , \tag{1.12}$$

and produces an element of $\mathcal{H}$. Therefore, an element of $\mathcal{H} \otimes \mathcal{H}^*$ constructed by using the map 1.11 can be seen as a function $\mathcal{H} \mapsto \mathcal{H}$. Therefore, $|\phi\rangle\langle\phi| \in \mathcal{M}\forall\ |\phi\rangle \in \mathcal{H}$. As both, $\mathcal{H}$ and $\mathcal{P}$ have the same dimension and are defined over the same field, it is true that $\mathcal{P} = \{f\left(|\phi\rangle\right) \forall\, |\phi\rangle \in \mathcal{H}\}$.

That is why one can also refer to elements of $\mathcal{H}$ as pure states. Moreover, this makes evident that the Hilbert space formalism alone would not be able to describe any mixed state. In order to include mixed states, one must appeal to the operator spaces formalism, i.e., the density operator formalism.

Up to now, nothing was said about the interpretation of the elements of a density operator. The objective of this section is presenting a reasonable justification for the interpretation of its diagonal elements as probabilities.

We begin the analysis by considering a quantum state $\rho \in \mathcal{M}$. Then consider a measurement operator $|\psi\rangle\langle\psi|$, where $|\psi\rangle \in \mathcal{H}$. After that measurement, the quantum system will be left in the state

$$\rho' = \frac{|\psi\rangle\langle\psi|^\dagger \rho |\psi\rangle\langle\psi|}{\mathrm{Tr}\left(|\psi\rangle\langle\psi|^\dagger \rho |\psi\rangle\langle\psi|\right)} = \frac{|\psi\rangle\langle\psi| \rho |\psi\rangle\langle\psi|}{\mathrm{Tr}\left(|\psi\rangle\langle\psi| \rho |\psi\rangle\langle\psi|\right)} = \frac{\langle\psi| \rho |\psi\rangle \, |\psi\rangle\langle\psi|}{\langle\psi| \rho |\psi\rangle \, \mathrm{Tr}\left(|\psi\rangle\langle\psi|\right)} = |\psi\rangle\langle\psi| \, . \tag{1.13}$$

Then, after that measurement the system will be in the pure state $|\psi\rangle$. The probability of that being the actual outcome of that measurement is then equal to the probability of ρ to be in the pure state $|\psi\rangle$. That probability is given by

$$p\left(|\phi\rangle\right) = \mathrm{Tr}\left(|\psi\rangle\langle\psi|^\dagger \rho |\psi\rangle\langle\psi|\right) = \langle\psi| \rho |\psi\rangle \, . \tag{1.14}$$

Now let the set $\{|e_i\rangle\langle e_j|\}_{i,j=1}^{\dim(\mathcal{H})}$ be an orthonormal basis for $\mathcal{M}$. In that basis, we can represent the density operator as

$$\rho = \sum_{i,j=1}^{\dim(\mathcal{H})} \alpha_{ij} \, |e_i\rangle\langle e_j| \, . \tag{1.15}$$

Its diagonal elements are

$$\langle e_n | \rho | e_n \rangle = \sum_{i,j=1}^{\dim(\mathcal{H})} \alpha_{ij} \langle e_n | e_i \rangle \langle e_j | e_n \rangle = \sum_{i,j=1}^{\dim(\mathcal{H})} \alpha_{ij} \delta_{in} \delta_{jn} = \alpha_{nn}. \tag{1.16}$$

Then, it is easy to see that the diagonal elements of ρ in that basis are given by $p(e_i)$. We call those diagonal elements populations as they represent how the state ρ is populated with respect to a given basis of $\mathcal{M}$. It is important to remark that populations are basis-dependent but still have physical significance. When one performs an experiment, the measurement apparatus usually defines a preferred basis.

The off diagonal elements of ρ are called coherences. These elements are related to the possibility of ρ to transit between two quantum states $|e_m\rangle$ and $|e_n\rangle$. The coherences are not a property that is exclusive of mixed nor entangled states. In fact, any pure state that is a superposition will present coherences. To see it, consider a pure state $|\phi\rangle \in \mathcal{H}$, we can start by writing it in some basis $\{|e_i\rangle\}_{i=1}^{\dim(\mathcal{H})}$:

$$|\phi\rangle = \sum_{i=1}^{\dim(\mathcal{H})} \beta_i |e_i\rangle . \tag{1.17}$$

Now we construct the associated density operator as

$$\rho = |\phi\rangle\langle\phi| = \sum_{i,j=1}^{\dim(\mathcal{H})} \beta_j^* \beta_i |e_i\rangle\langle e_j| , \tag{1.18}$$

and decompose it as

$$\rho = \sum_{i=1}^{\dim(\mathcal{H})} |\beta_i|^2 \,|e_i\rangle\langle e_i| + \sum_{i,j=1;i>j}^{\dim(\mathcal{H})} \beta_j^* \beta_i \,|e_i\rangle\langle e_j| + \sum_{i,j=1;i<j}^{\dim(\mathcal{H})} \beta_j^* \beta_i \,|e_i\rangle\langle e_j| \,. \tag{1.19}$$

Then, ρ will not present coherences in the given basis if and only if

$$\beta_j \beta_i = 0 \forall i \neq j, \tag{1.20}$$

in other words, if and only if it does not present any superposition in that basis.

Coherences are an interesting property of quantum systems, in the sense that if a quantum state does not present coherences, it will not present entanglement nor superpositions. Therefore a classical system can efficiently simulate its behavior. Moreover, quantum coherences are volatile, in the sense that if a quantum system is left in contact with the environment, it will quickly suffer decoherence, i.e., its off-diagonal elements will evolve towards zero. After complete decoherence, the system will become completely mixed, and we say that we have lost information about its quantum state. Decoherence is one of the main obstacles when one thinks about developing new technologies that demand quantum systems to behave in a very controlled manner.

1.5 Entanglement

Among all subtleties of quantum mechanics, the possibility of quantum entanglement is by far one of the most amazing possibilities because it is an exclusively quantum resource[1].

[1] Classical systems can present entanglement but only between degrees of freedom that are not spatially separated. For more on this, see [14, 15, 16, 17].

In this section, the idea of entanglement will be presented for the case of bipartite pure states as the treatment of other cases would be out of the scope of this text. For a complete treatment of entanglement, the reader is referred to [11].

We begin by presenting a common example of entanglement, which is the one that can happen between two qubits, i.e., two-level systems. To represent a qubit we use the Hilbert space $\mathcal{H} = \mathrm{span}\left\{|0\rangle, |1\rangle\right\}$. The most general pure qubit state is

$$|\psi\rangle = \alpha_0 |0\rangle + \alpha_1 |1\rangle \,; \alpha_0, \alpha_1 \in \mathbb{C}, \tag{1.21}$$

such that

$$|\alpha_0|^2 + |\alpha_1|^2 = 1. \tag{1.22}$$

It is then simple to see that one can alternatively write such state as

$$|\psi\rangle = \sin\frac{\theta}{2} |0\rangle + e^{i\phi} \cos\frac{\theta}{2} |1\rangle \,; \theta \in [0, \pi] \,, \phi \in [0, 2\pi[\,, \tag{1.23}$$

where a global phase factor was neglected without loss of generality. A global phase factor produces no observable effects. By writing the qubit state in that form, it becomes evident that any quantum state of $\mathcal{H}$ is in the surface of the so-called Bloch sphere, as presented in figure 1.1. Therefore, a single 3-vector is sufficient to completely characterize the state. The Hilbert space for a two qubit system is $\mathcal{H}_2 = \mathcal{H} \otimes \mathcal{H}$. By looking at the Bloch sphere, one might be tempted to think that any 2-qubit state can be represented as the product of two qubit states, as depicted in figure 1.2. But this is note the case. To understand why, we can

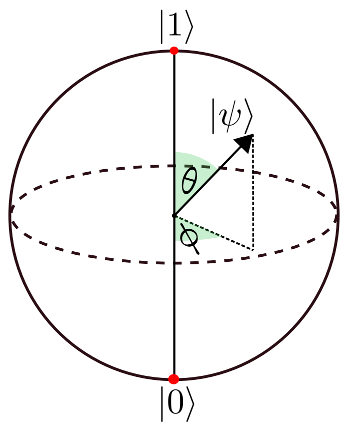

Figure 1.1: A Bloch sphere and a qubit state represented.

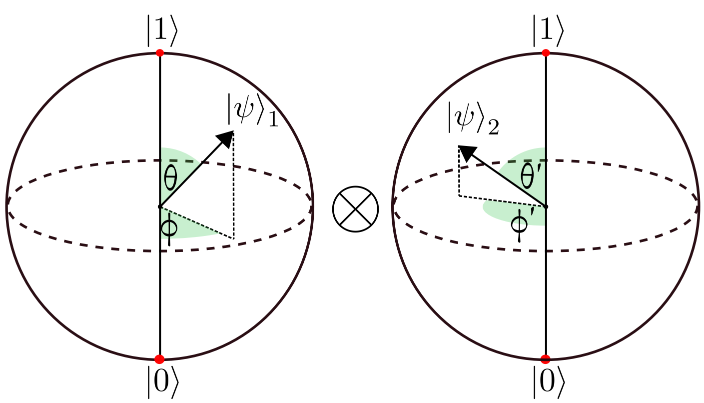

Figure 1.2: A 2-qubit state formed by the product of two qubit states.

write down the depicted 2-qubit state:

$$|\psi\rangle_1 |\psi\rangle_2 = \sin\frac{\theta_1}{2}\sin\frac{\theta_2}{2}|0\rangle|0\rangle + \sin\frac{\theta_1}{2}\cos\frac{\theta_2}{2}e^{i\phi_2}|0\rangle|1\rangle + \sin\frac{\theta_2}{2}\cos\frac{\theta_1}{2}e^{i\phi_1}|1\rangle|0\rangle + \\ + e^{i(\phi_1+\phi_2)}\cos\frac{\theta_1}{2}\cos\frac{\theta_2}{2}|1\rangle|1\rangle\,; \theta_1,\theta_2 \in [0,\pi]\,, \phi_1,\phi_2 \in [0,2\pi[\,. \tag{1.24}$$

Notice that to completely describe such state one needs to provide four parameters: $\theta_1, \theta_2, \phi_1, \phi_2$. With that result in mind, we proceed to write the most general 2-qubit state, i.e., the most general $|\psi\rangle \in \mathcal{H}_2$:

$$|\psi\rangle = \alpha_{00}\,|0\rangle\,|0\rangle + \alpha_{01}\,|0\rangle\,|1\rangle + \alpha_{10}\,|1\rangle\,|0\rangle + \alpha_{11}\,|1\rangle\,|1\rangle\,; \alpha_{ij} \in \mathbb{C}, \tag{1.25}$$

such that

$$|\alpha_{00}|^2 + |\alpha_{01}|^2 + |\alpha_{10}|^2 + |\alpha_{11}|^2 = 1. \tag{1.26}$$

Therefore, the most general 2-qubit state is described by 8 real parameters, but one of them, the global phase, is irrelevant, and another of them is constrained due to normalization. That equates to 6 free parameters wherein the product state of two qubits we have 4 parameters. That is an indication that there are 2-qubit states that cannot be represented as a product of two qubit states. Quantum states that cannot be separated as a product of two well-defined quantum states are called entangled states. An example of a non-entangled 2-qubit state is

$$|\psi\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle\,|1\rangle + |1\rangle\,|1\rangle\right). \tag{1.27}$$

That state can be decomposed as the following product state:

$$|\psi\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle + |1\rangle\right)|1\rangle\,, \tag{1.28}$$

meaning that one of the qubits is in the state $|1\rangle$ and the other is in the state $\frac{1}{\sqrt{2}}\left(|0\rangle + |1\rangle\right)$. An example of entangled state for the 2-qubit system is one of the Bell states[10],

$$\left|\phi^+\right\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle\,|0\rangle + |1\rangle\,|1\rangle\right). \tag{1.29}$$

It is not possible to write it as a product state, therefore, despite being a 2-qubit state, none of the two qubits have a well-defined quantum state by themselves. Quantum entanglement is a global property of the system.

Despite the simple system used for the example, the entanglement concept is valid for any composite quantum system. We remark that knowing if a pure quantum state is separable, i.e., if it can be represented as a convex sum of products of quantum states, is nowhere near to be trivial[11]. Luckily, for the case of a pure quantum state, it is possible to determine if the quantum state of the full system can be written as the product of two quantum states of its parts.

Let $\mathcal{H}$ be a Hilbert space and $\mathcal{M}$ be the space of operators $\mathcal{H} \mapsto \mathcal{H}$. Then consider two Hilbert spaces $\mathcal{A}$ and $\mathcal{B} \subset \mathcal{H}$ such that $\mathcal{H} = \mathcal{A} \otimes \mathcal{B}$. Additionally, let $\dim(\mathcal{A}) > \dim(\mathcal{B})$. For a given pure quantum state $|\psi\rangle \in \mathcal{H}$, one defines the entanglement entropy[2] as

$$E(|\Psi\rangle) \equiv S(\rho_A) = -\operatorname{Tr}(\rho_A \ln \rho_A), \tag{1.30}$$

where

$$\rho_A = \operatorname{Tr}_B |\Psi\rangle\langle\Psi|. \tag{1.31}$$

The entanglement entropy is bounded by $[0, \ln \dim(\mathcal{B})]$.

As an example, consider the Bell state in equation 1.29. Let A be the first qubit and B be

[2] If the reader feels confused with the several entropies, we refer the reader to appendix D.

the second one. By tracing out the second qubit we get

$$
\begin{aligned}
\rho_A &= \mathrm{Tr}_{\mathcal{B}}\left[\frac{1}{2}\left(|0\rangle\,|0\rangle\,\langle 0|\,\langle 0| + |0\rangle\,|0\rangle\,\langle 1|\,\langle 1| + |1\rangle\,|1\rangle\,\langle 0|\,\langle 0| + |1\rangle\,|1\rangle\,\langle 1|\,\langle 1|\right)\right] \\
&= \frac{1}{2}\sum_{i=0}^{1}\mathbf{1}\otimes\langle i|\left(|0\rangle\,|0\rangle\,\langle 0|\,\langle 0| + |0\rangle\,|0\rangle\,\langle 1|\,\langle 1| +\right. \\
&\left. + |1\rangle\,|1\rangle\,\langle 0|\,\langle 0| + |1\rangle\,|1\rangle\,\langle 1|\,\langle 1|\right)\mathbf{1}\otimes|i\rangle \\
&= \frac{1}{2}\left(|0\rangle\!\langle 0| + |1\rangle\!\langle 1|\right). \qquad (1.32)
\end{aligned}
$$

It means that the qubit A could be in states $|0\rangle$ or $|1\rangle$ with equal chance. By ignoring the qubit B, we are left with a completely random state, i.e., all possible outcomes are equiprobable. The entanglement entropy between A and B is then

$$
S\left(\rho_A\right) = -\,\mathrm{Tr}\left(\rho_A \ln \rho_A\right) = -\sum_{i=1}^{2} p_i \ln p_i, \qquad (1.33)
$$

where p_i are the eigenvalues of ρ_A. Continuing the calculation,

$$
S\left(\rho_A\right) = -2\frac{1}{2}\ln\left(\frac{1}{2}\right) = \ln 2. \qquad (1.34)
$$

Then, A is maximally entangled with B as $\dim\left(\mathcal{A}\right) = 2$.

If two systems A and B are maximally entangled, then, by tracing out one of them, say B, one is left with system A in a completely random state.

As another example, consider the quantum state of equation 1.28. By tracing out system

B, system A is left in the state

$$\rho_A = \frac{1}{2} \left(|0\rangle\langle 0| + |0\rangle\langle 1| + |1\rangle\langle 0| + |1\rangle\langle 1| \right). \tag{1.35}$$

That density operator has only one non-zero eigenvalue which is equal to 1, therefore, its entanglement entropy is zero. It means that there is no entanglement between system A and B.

If there is no entanglement between A and B, then, by tracing out one of them, say B, one is left with system A in a pure state. It means that there will be no correlations between outcomes of experiments done in A and B.

Two quantum systems can be entangled but not maximally entangled. In that case, the entanglement entropy will be neither maximum nor minimum. For these cases, tracing out one of the systems will leave the other in a quantum state that is not entirely random.

A crucial property is that when a quantum system in a pure state is biparted, i.e., divided into two parts as we did, the entanglement entropy is the same for both parts, that is,

$$S\left(\rho, \mathcal{A}\right) = S\left(\rho, \mathcal{B}\right). \tag{1.36}$$

The proof for this property is based on Schmidt decomposition theorem and can be found as part of theorem 11.8 of [10].

1.6 A brief introduction to Quantum Information

We will begin the discussion with a qualitative introduction to what "is" information for then introduce its mathematical definition.

Suppose that someone challenged you on the hangman game. The person who challenged you then draw the standard illustration of the game, which is on figure 1.3. You have

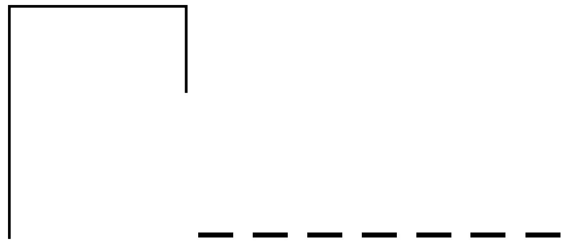

Figure 1.3: The beginning of a hangman game.

to guess an English word that is 7 letters long. Considering only combinatorial analysis, the set of possible guesses is

$$\Omega(\varnothing) = \{all\ 7\ letters\ anagrams.\} \tag{1.37}$$

which has $|\Omega(\varnothing)| = 24^7 = 4586471424$ elements. The $\varnothing$ means that no correct guesses were done. Then consider that you correctly guessed letters h, i, p, s, the board now looks like the illustration in figure 1.4.

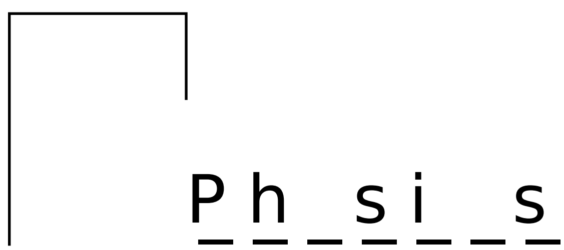

Figure 1.4: The climax of a hangman game.

At that point, there are only 2 characters left to guess. Therefore, the set of possible words is

$$\Omega(h, i, p, s) = \{all\ 2\ letters\ anagrams.\}, \tag{1.38}$$

which should fill the empty spaces in figure 1.4. There are $|\Omega(h,i,p,s)| = 24^2 = 576$ elements now. A lot less than before, i.e., $|\Omega(h,i,p,s)| < |\Omega(\varnothing)|$. It might be obvious but the more letters you guess, the less possibilities are left in the set of possible guesses.

Once you guess the full word, the board will look like the illustration in figure 1.5. At the

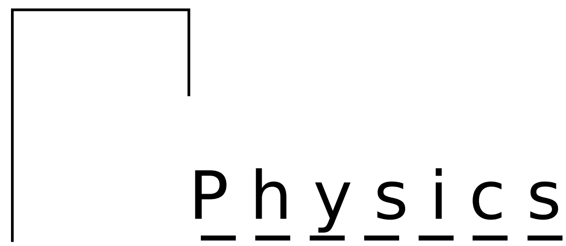

Figure 1.5: The end of a hangman game.

end of a hangman game, the set of possible guesses will be reduced to one element only: the correct word. Therefore, a hangman game ends when $\Omega = \{\textit{the correct word}\}$, or equivalently $|\Omega| = 1$.

A hangman game has a lot to do with information. From a more abstract point of view, you begin a game with no information about the word you will guess at the end which leads to a huge set of possible guesses. After each guess, be it correct or wrong, the set of possible guesses will for sure dwindle. That happens because you are gathering information about that word. From that example it is easy to see that from an abstract point of view, information has to do with the set of possible guesses. It is then reasonable to qualitatively define information as anything that eliminates elements from the set of possible guesses.

The qualitative definition given in the last paragraph is not the best one could get as information does not exist by itself. One always say to have information about something, that is, information is always with respect to something. In the case of the hangman game, it was information about the unknown word. Therefore, a refined, but still qualitative definition is that information about something is anything that eliminates elements from the set of possible guesses about that thing.

That refined definition is important because it enforces the connection between information and correlation. For instance, during a hangman game if someone had told you that the temperature outside is $38°C$, it would had not eliminated any word from the set of possible guesses as the climate has no correlation with the word chosen by the challenger.[3] Then, as the climate is uncorrelated to the hangman game, getting information about the climate will not lead to any information about the unknown word. Therefore, to eliminate elements from the set of possible guesses, there must be correlation between them. In other words, information is always encoded in correlations about two things.

In a more precise way, Shannon presents a measure of how big the set of possible outcomes, which we were referring to as "set of possible guesses", is. That measure is defined in [18] for a set Ω of possible outcomes with probabilities $\{p_i\}_{i=1}^{n}$ as:

$$H\left(p_1,\ldots,p_n\right) = -K\sum_{i=1}^{n} p_i \ln p_i, \tag{1.39}$$

where K is a positive constant and the p_i are the probabilities for the outcome $i \in \Omega$. Therefore, H can be understood to be a measure of our ignorance about the outcome of an event.

In Quantum mechanics, the density operator describes the probability of each outcome to happen, therefore, one can calculate its Shannon entropy as

$$H\left(\rho\right) = -K\,\mathrm{Tr}\,\rho\ln\rho. \tag{1.40}$$

[3]That is supposing that the challenger is not even aware of the climate.

By taking $K = 1$ one gets the so-called Von Neumann entropy. Therefore, one can understand the Von Neumann entropy of a quantum state ρ as a measure of ignorance about the exact quantum state of a system.

The relation between Quantum Information and Thermodynamics will be delayed until after chapter 2.

2

Thermodynamics and Statistical Mechanics

During the early discussions on Black Hole entropy, thermodynamics played a fundamental role because according to its laws, everything that has a thermodynamic entropy is also expected to have a temperature and therefore, to emit thermal radiation. During the early discussions, this was a significant motivation for further investigations on the issue, which led to Hawking's remarkable result on Black Hole radiation. This chapter is dedicated to briefly review the laws of thermodynamics and the formalism of statistical mechanics.

2.1 The laws of thermodynamics

Born as an empirical theory, thermodynamics as we know it today started as a study about the conversion of heat into work in 1824 with [19]. A later development of that work resulted in what we know today as the second law of thermodynamics, formulated for the first time in both [20] and [21], where the latter presented a formulation of the first law too. The third law came only in 1912 with Nernst's statement, as pointed in [22]. Ironically, the zeroth law was the last one to be stated. It happened only in 1939 in [23]. It is important to remark that the idea of thermal equilibrium, which is behind the zeroth law was already present in many texts but not with the status of a law.

Despite being entirely built on an empirical basis, the power of thermodynamics relies on its generality. The laws of thermodynamics do not depend on the particular constituents of the system one wishes to study but rather on general considerations about heat transfer and its capacity of being converted into work.

There were also efforts trying to study thermodynamics in a mathematically rigorous way by giving it an axiomatic treatment where the laws of thermodynamics have the status of theorems. The first attempt occurred in 1909 on [24] with a geometrical approach to thermodynamics. Further development of the formalism contributed to the foundation of a new topic called geometrothermodynamics [25], where the full machinery of differential geometry is used to describe thermodynamics. We now begin presenting the laws of thermodynamics and their consequences.

Definition 2.1.1 (Thermodynamic equilibrium). (Adapted from [26])An isolated system[1] is said to be in thermodynamic equilibrium when there are no net macroscopic flows of energy or matter between its parts.

A thermodynamic equilibrium state of a system is characterized by the values of a set of state variables, that is, a set of values of macroscopic quantities that depend only on the current thermodynamic equilibrium state of the system and not on the path it followed to achieve it. Moreover, any relation between state variables is called a state function.

Zeroth law of thermodynamics 2.1.1. (As is in [27]) If two systems, A and B, are separately in thermodynamic equilibrium with a third system, C, then they are also in thermodynamic equilibrium with one another.

A consequence of the zeroth law is the possibility to define a state function Θ_M for a system M, which we recognize as the thermodynamic temperature, as follows: Let two systems A and B be in thermodynamic equilibrium states described by the sets $\mathcal{A}$ and $\mathcal{B}$ of values of state variables, respectively. If A and B are in thermodynamic equilibrium with one another, they have the same temperature,

$$\Theta_A\left(\mathcal{A}\right) = \Theta_B\left(\mathcal{B}\right). \tag{2.1}$$

The sets $\mathcal{A}$ of values of state variables satisfying the equation of state $\Theta_A\left(\mathcal{A}\right) = \Theta$, where $\Theta \in \mathbb{R}$ is a constant, compose the isotherms of A.

[1]An isolated system that can not exchange matter nor energy with the environment.

Several different scales can be used to measure the temperature Θ. Throughout this text, we will use the Kelvin scale, and to emphasize that, we will use T for the temperature.

First law of thermodynamics 2.1.1. (As is in [27]) The amount of work required to change the state of an otherwise adiabatically isolated[2] system depends only on the initial and final states, and not on the means by which the work is performed, or on the intermediate stages through which the system passes.

From that law, one can recover a statement of energy conservation. Let a system A transform adiabatically (only exchanging energy in the form of mechanical work) from a thermodynamic equilibrium state, $\mathcal{A}_i$, to another one, $\mathcal{A}_f$. Denoting by ΔW the amount of work done on the system, one can mathematically state the first law as

$$\Delta W = E\left(\mathcal{A}_f\right) - E\left(\mathcal{A}_i\right) = \Delta E, \tag{2.2}$$

where E is a state function called the internal energy of the system. By repeating the same transformation, $\mathcal{A}_i \to \mathcal{A}_f$ without assuming that it happens adiabatically, one would find that

$$\Delta E \neq \Delta W, \tag{2.3}$$

that is, the internal energy variation is not equal to the work done on the system, motivating the definition of a new quantity,

$$\Delta Q = \Delta E - \Delta W, \tag{2.4}$$

[2] An adiabatically isolated can only exchange energy with the environment in the form of mechanical work.

where ΔQ is the heat intake of the system from the environment. In such transformations, both, ΔQ and ΔW are not separately functions of states because they depend on the particular process and not only on the initial and final thermodynamic equilibrium states. A mathematical way of phrasing that is

$$\mathrm{d}E = đQ + đW, \tag{2.5}$$

where $đ$ is an inexact differential that cannot always be obtained by differentiation.

One can typically write $đW$ as

$$đW = \sum_i F_i \, \mathrm{d}x_i \, , \tag{2.6}$$

where x_i is a state variable, understood in this context as a generalized coordinate, and F_i is the generalized conjugate force associated with the generalized displacement $\mathrm{d}x_i$. In the context of thermodynamics, both, the generalized coordinate and the associated generalized conjugate force are pairs of state variables.

A remarkable consequence of the first law is forbidding the existence of engines that can produce work without consuming energy. However, the first law does not constrain the efficiency of a thermal machine. It is the second law that does.

Second law of thermodynamics 2.1.1. (Adapted from [27]) (Kelvin's statement). No cyclic process is possible whose sole result is the complete conversion of heat into work.

There are two significant consequences of this law, and both come in the form of theorems whose proofs can be found in[27]. The first one makes use of a specific kind of engine defined next.

Definition 2.1.2 (Carnot engine). (As is from [27])A Carnot engine is any engine that is reversible[3], runs in a cycle, with all of its heat exchanges taking place at a source temperature T_H, and a sink temperature T_C.

Theorem 2.1.1 (Carnot's theorem). (As is from [27])No engine operating between two reservoirs (at temperatures T_H and T_C) is more efficient[4] than a Carnot engine operating between them.

Such theorem is relevant because it is necessary to derive Clausius's theorem 2.1.2. Another consequence is that it puts boundaries on the efficiency of thermal machines.

Theorem 2.1.2 (Clausius's theorem). (As is from [27])For any cyclic transformation (reversible or not), $\oint \frac{đQ}{T} \leq 0$, where $đQ$ is the heat increment supplied to the system at temperature T.

Let $\mathcal{A}$ and $\tilde{\mathcal{A}}$ be thermodynamic equilibrium states of the same system. Consider a reversible cycle $\mathcal{A} \xrightarrow{(1)} \tilde{\mathcal{A}} \xrightarrow{(2)} \mathcal{A}$. For clarity, we will use the subscript rev to indicate that the heat exchange is due to a reversible process, and the superscript in parentheses to indicate what part of the process we are referring to. Under that reversible cycle[5],

$$\oint \frac{đQ_{rev}}{T} = \int_{\mathcal{A}}^{\tilde{\mathcal{A}}} \frac{đQ_{rev}^{(1)}}{T^{(1)}} + \int_{\tilde{\mathcal{A}}}^{\mathcal{A}} \frac{đQ_{rev}^{(2)}}{T^{(2)}} = 0. \tag{2.7}$$

Therefore,

$$\int_{\mathcal{A}}^{\tilde{\mathcal{A}}} \frac{đQ_{rev}^{(1)}}{T^{(1)}} = \int_{\mathcal{A}}^{\tilde{\mathcal{A}}} \frac{đQ_{rev}^{(2)}}{T^{(2)}}, \tag{2.8}$$

[3]A reversible engine operates only by reversible processes

[4]In this context, we are concerning the efficiency with which the engine converts heat into work.

[5]Any reversible process is a quasi-static process. Therefore, the system in question has well-defined thermodynamic equilibrium state during the whole process.

implying that the result of the integrals depends only on the thermodynamic equilibrium states $\mathcal{A}$ and $\tilde{\mathcal{A}}$ but not on the chosen reversible process, characterizing it as a function of state. Motivating the definition of thermodynamic entropy, up to an integration constant, as

Definition 2.1.3 (Thermodynamic entropy). The variation of thermodynamic entropy when a system goes through a reversible process from an equilibrium state, $\mathcal{A}$, to another one, $\mathcal{B}$, is

$$S\left(\mathcal{B}\right)-S\left(\mathcal{A}\right)\equiv\int_{\mathcal{A}}^{\mathcal{B}}\frac{đQ_{rev}}{T}. \tag{2.9}$$

For reversible processes it is then possible to recognize $đQ = T\,\mathrm{d}S$ and hence write the energy conservation statement (2.5) as

$$\mathrm{d}E=T\,\mathrm{d}S+\sum_{i}F_i\,\mathrm{d}x_i\,, \tag{2.10}$$

where we used equation (2.6). Note that it is a relation between functions of state, therefore it is valid even when the process under consideration is not reversible.

Furthermore, consider an arbitrary process, $\mathcal{A}\to\mathcal{B}$, followed by a reversible process, $\mathcal{B}\to\mathcal{A}$. According to theorem 2.1.2,

$$\oint\frac{đQ}{T}=\int_{\mathcal{A}}^{\mathcal{B}}\frac{đQ}{T}+\int_{\mathcal{B}}^{\mathcal{A}}\frac{đQ_{rev}}{T}\leq 0. \tag{2.11}$$

Using the definition 2.1.3 of thermodynamic entropy, one can write it as

$$\int_{\mathcal{A}}^{\mathcal{B}}\frac{đQ}{T}\leq S\left(\mathcal{B}\right)-S\left(\mathcal{A}\right). \tag{2.12}$$

In differential form,

$$dS \geq \frac{\text{đ}Q}{T}, \tag{2.13}$$

for any transformation. The equality holds only in reversible processes.

Finally, considering an adiabatically isolated system, one gets $\text{đ}Q = 0$, as an adiabatically isolated can only exchange energy with another system in the form of mechanical work. In that situation, equation (2.13) leads to the famous conclusion:

$$dS \geq 0, \tag{2.14}$$

Where the equality only happens when the underlying thermodynamic process is reversible. In other words, the thermodynamic entropy of an adiabatically isolated system either increases or stays constant.

As an example of how the thermodynamic entropy of an adiabatically isolated system can increase, consider a closed system A in a thermodynamic equilibrium state, $\mathcal{A}$, at temperature is T_A. Then consider another closed system B in another thermodynamic equilibrium state, $\mathcal{B}$, at temperature $T_B < T_A$. Let A and B be in contact. The complete $A + B$ system is adiabatically isolated as there is nothing else being considered other than A and B. Due to putting in contact two systems that are not in thermodynamic equilibrium with one another, net macroscopic flows of energy and matter will arise taking the system into a transitory non-equilibrium state which will last until the macroscopic flows of energy and matter cease, leaving the complete system in an equilibrium state $\mathcal{C}$. The temperature of both A and B is equal to $T_B < T_C < T_A$. The entropy change in the complete system during the thermalization process is

$$\Delta S = S_A(\mathcal{C}) + S_B(\mathcal{C}) - S_A(\mathcal{A}) - S_B(\mathcal{B}), \tag{2.15}$$

where $S_M(\mathcal{M})$ is the thermodynamic entropy of system M that is in a thermodynamic equilibrium state[6] $\mathcal{M}$.

It can be shown that the second law of thermodynamics implies that heat always flows towards the lower temperature system, then, any heat coming into system B must come from system A, implying $đQ_B = -đQ_A$. As the temperature of system B is smaller than that of system A, $đQ_B > 0$ and $đQ_A < 0$ but both have the same absolute value. With that in mind, one concludes that for any reversible process connecting both thermal states $\mathcal{A}$ and $\mathcal{B}$ to the state of mutual thermodynamic equilibrium $\mathcal{C}$,

$$\frac{đQ_B}{T_B} > \left|\frac{đQ_A}{T_A}\right|, \tag{2.16}$$

as $T_B < T_A$ during the thermalization. Therefore,

$$\mathrm{d}S \geq -\left|\frac{\mathrm{d}Q_A}{T_A}\right| + \frac{\mathrm{d}Q_B}{T_B} > 0. \tag{2.17}$$

Which finally leads to a net increase in the full system entropy, as $\mathrm{d}S > 0$ during such process. As the entropy is a function of state, it does not depend on the particular process that connects thermal states. Therefore $\Delta S > 0$ during the thermalization process.

Third law of thermodynamics 2.1.1. (As is in [28]) The change in entropy which occurs

[6] Also known as thermal state.

when a homogeneous system undergoes an isothermal reversible process approaches zero as the temperature approaches absolute zero.

As a consequence of the third law, one can show that the entropy tends to a constant value, which can in principle depend on the system, as the temperature tends to zero, that is:

$$\lim_{T \to 0} \left. \frac{\partial S}{\partial X} \right|_T = 0, \tag{2.18}$$

for any state variable X. Then, it is easy to see that as T tends to zero, the entropy tends to a constant as its variation with respect to any state variable tends to zero. Therefore, one concludes that

$$\lim_{T \to 0} S\left(T, \mathcal{X}\right) = S_M, \tag{2.19}$$

where $\mathcal{X}$ is the set of state variables associated with the system M, and S_M is the value of its entropy when the temperature tends to absolute zero. This equivalent to the statement [29]:

> the entropy of every pure, perfectly crystalline substance approaches the same value as the temperature approaches zero

Then, one can set $S_M = S_0$, where S_0 is the entropy of any pure, perfectly crystalline substance M. For systems that are not of this kind, some residual entropy is expected to persist even when the temperature approaches absolute zero.

With that in mind, it is possible to define $S_0 = 0$, allowing one to make sense of the thermodynamic entropy by itself, at least for perfectly crystalline substances, and not only of its variation. For those substances, one can choose a standard state $\mathcal{R}\left(T = 0\right)$, which is at zero absolute temperature and has zero thermodynamic entropy as a reference and calculate the entropies:

$$S(\mathcal{A}) = \int_{\mathcal{R}(T=0)}^{\mathcal{A}} \frac{\text{đ}Q_{rev}}{T}. \tag{2.20}$$

For non perfectly crystalline substances, statistical mechanics sheds some light on the value of the entropy.

Additionally, it can be shown that the third law implies that

1. Heat capacities must vanish as $T \to 0$, so, $\lim_{T\to 0} C_X(T) = 0$.
2. It is impossible to cool any system to absolute zero temperature in a finite number of steps.

for a derivation of those consequences, one can refer to [27].

2.2 Statistical Mechanics

By the time the laws of equilibrium thermodynamics were being stated as we know them today, statistical mechanics began to take form. Maxwell formulated the first statistical law in physics in 1860 in the articles [30] and [31]. The term statistical mechanics was coined by Gibbs in his article [32] and he also wrote a book [33] that gave statistical mechanics the form we know today. Gibb's wrote that book [33] in 1902, and Quantum Mechanics was not fully developed yet. Therefore, his approach was supposed to be valid only for classical mechanics. However, the framework was adapted to Quantum Mechanics in 1938 by Tolman in his book [34], and it still holds to this day.

For the discussion presented the present text, it suffices to understand quantum statistical mechanics. Hence, only this formulation will be presented. It is important to note, however, that while the formulation of classical statistical mechanics resembles the quantum formulation, there are several different mathematical details.

Definition 2.2.1 (Microstate). A microstate of a quantum system is a pure state $\rho \in \mathcal{M}$, where $\mathcal{M}$ is the state space of the system.

Definition 2.2.2 (Macrostate). A macrostate $M(\mathcal{X})$ of a system is defined by a finite set of values of distinct state variables $\mathcal{X} = (X_1, \ldots, X_N)$.

Definition 2.2.3 (Ensemble). An ensemble is the set of microstates corresponding to a given macrostate.

2.2.1 Microcanonical Ensemble

Generally, one can determine the values of the state variables of a system in thermodynamic equilibrium by making measurements. Throughout this section, we will use $\mathcal{X}$ as that set of values of state variables. In the situation where one only knows value of the energy E, the occupied volume V and the number of particles N in the system, the macrostates are $M(E, V, N)$. The set of microstates corresponding to this macrostate forms an ensemble called the microcanonical ensemble.

The postulate of equal a priori equilibrium probabilities says that every microstate within the ensemble is equally likely to be accessed by a system in thermodynamic equilibrium. Therefore, the density operator for the microcanonical ensemble in energy eigenbasis is

$$\rho(E, V, N) = \frac{1}{\Omega(E, V, N)} \sum_n \delta(E_n - E) |n\rangle\langle n|, \tag{2.21}$$

where E_n is the eigenvalue of the energy eigenstate $|n\rangle$, and $\Omega\left(E, V, N\right)$ is the number of microstates compatible with the macrostate $M\left(E, V, N\right)$, which in this situation is the number of N particle eigenstates contained within a volume V such that the expected energy is E.

One can then proceed and calculate the Boltzmann entropy of the microcanonical ensemble as

$$S\left(E, V, N\right) = K_B \ln \Omega\left(E, V, N\right). \tag{2.22}$$

The equivalence between the Boltzmann entropy and the thermodynamic entropy follows from the equivalence between statistical mechanics and thermodynamics presented in the next section.

2.2.2 Equivalence with equilibrium thermodynamics

In this section the justification for the equivalence between statistical mechanics and equilibrium thermodynamics will be presented, using [27], [34] and [35] as references.

We will begin by giving a reasoning to explain the zeroth law. Consider two isolated systems separately in thermodynamic equilibrium as illustrated in figure 2.1. Consider that they are in microcanonical ensembles described by the macrostates $M\left(E_1, N_1, V_1\right)$ and $M\left(E_2, N_2, V_2\right)$. Let the systems interact exchanging energy but not work. The system will

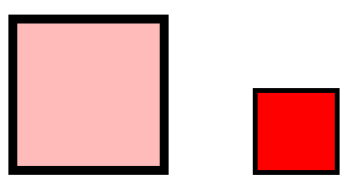

Figure 2.1: Two systems in separate thermodynamic equilibrium. The system on the left is at a lower temperature than the system on the right. The black borders indicate insulating walls

go through a transient non-equilibrium stage and after a sufficient amount of time it will reach thermodynamic equilibrium. The composed system in thermodynamic equilibrium produces a third system in the macrostate $M\left(E, N_1+N_2, V_1+V_2\right)$, where $E=E_1+E_2$, illustrated in figure 2.2.

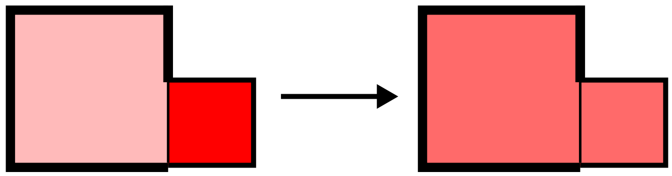

Figure 2.2: The two systems interact through the thinner wall, which allows energy exchange but is rigid and immovable, preventing mechanical work. One should understand time as passing from the left to the right. After the thermalization process the whole system has the same color, indicating that they are at the same temperature.

According to the postulate of equal a priori equilibrium probabilities, all possible microstates are equally likely to happen. Therefore, the full system has no preferred partition for the energy among the subsystems. The only constraint is that the total energy must sum to E. Therefore, to calculate the number of accessible microstates of the joint system one must consider all possibilities of energy partitioning,

$$\Omega\left(E\right)=\int_0^E \mathrm{d}\varepsilon\, \Omega_1\left(\varepsilon, V_1, N_1\right)\Omega_2\left(E-\varepsilon, V_2, N_2\right)=\int_0^E \mathrm{d}\varepsilon\, e^{\ln\Omega_1\left(\varepsilon, V_1, N_1\right)+\ln\Omega_2\left(E-\varepsilon, V_2, N_2\right)}, \tag{2.23}$$

where Ω_1 and Ω_2 are the number of accessible microstates in systems 1 and 2, given their volume, internal energy and number of particles. The number of accessible microstates in each subsystem is exponential in the number of particles as each subsystem is composed by indistinguishable non-correlated particles. Hence, the above expression for the number of microstates can be written as

$$\Omega\left(E\right)=\int_{0}^{E}\mathrm{d}\varepsilon\, e^{N\left[\frac{N_1}{N}\ln\Omega_1\left(\varepsilon,V_1,1\right)+\frac{N_2}{N}\ln\Omega_2\left(E-\varepsilon,V_2,1\right)\right]}, \tag{2.24}$$

where $N = N_1 + N_2$. This integral can be estimated using the method of steepest descent with a relative error of order N^{-1}. Typically, one consider systems where $N \geq 10^{23}$ and in those cases the error is negligible. The dominant contribution to the number of accessible microstates comes from a particular value of ε that maximizes the exponent in the integrand. To obtain such value one can solve

$$\frac{\partial}{\partial\varepsilon}\left[\frac{N_1}{N}\ln\Omega_1\left(\varepsilon,V_1,1\right)+\frac{N_2}{N}\ln\Omega_2\left(E-\varepsilon,V_2,1\right)\right]=0 \tag{2.25}$$

for ε as the subsystems can exchange energy but not work or particles. The actual solution is not of interest for this discussion. Instead one can recognize the Boltzmann entropy in the condition 2.25 and write it as

$$\frac{\partial}{\partial\varepsilon}\left[S_1\left(\varepsilon,V_1,N_1\right)+S_2\left(E-\varepsilon,V_2,N_2\right)\right]=0, \tag{2.26}$$

where S_1 and S_2 are the Boltzmann entropies of systems 1 and 2 respectively. We will make a brief pause in the calculations to highlight some important consequences of what we had just done.

We begun trying to calculate the number of accessible microstates for a system in a macrostate $M\left(E, N_1+N_2, V_1+V_2\right)$. That system is composed by two smaller systems and at the beginning of the calculations we had no clue about what would be their macrostates given that the joint system is in $M\left(E, N_1 + N_2, V_1 + V_2\right)$. We then had to consider every single energy partition compatible with $E = E_1 + E_2$. By massaging equation (2.23) we got to (2.24) and concluded that the steepest descent approximation can be trusted as the relative error is of order N^{-1} and usually $N \geq 10^{23}$. Therefore, despite the fact that each microstate is equally probable, there are exponentially more microstates close to a particular partition of energy that is given by ε and $E - \varepsilon$. A physical consequence is that after reaching thermodynamic equilibrium, the joint system will spend much more time with approximately that partition of energy. The value of ε is such that equation (2.26) is satisfied, therefore, it is such that the entropy of the full system is extremized. Then, we conclude that after reaching thermodynamic equilibrium, the system will spend much more time in a configuration that extremizes the entropy. We can therefore recognize Boltzmann entropy as a thermodynamic entropy given that both have compatible properties. We now go back to the calculations.

Defining $\tilde{\varepsilon} = E - \varepsilon$ and working the derivatives, one gets

$$\frac{\partial}{\partial \varepsilon} S_1\left(\varepsilon, V_1, N_1\right) = \frac{\partial}{\partial \tilde{\varepsilon}} S_2\left(\tilde{\varepsilon}, V_2, N_2\right), \tag{2.27}$$

which can be written as

$$\left.\frac{\partial S_1\left(\varepsilon, \mathcal{X}_1\right)}{\partial \varepsilon}\right|_{\mathcal{X}_1} = \left.\frac{\partial S_2\left(\tilde{\varepsilon}, \mathcal{X}_2\right)}{\partial \tilde{\varepsilon}}\right|_{\mathcal{X}_2}, \tag{2.28}$$

where $\mathcal{X}_1 = \left(V_1, N_1\right)$ and $\mathcal{X}_2 = \left(V_2, N_2\right)$. Equation 2.28 establishes a relation between two functions of state of systems in thermodynamic equilibrium with one another, allowing one to identify the thermodynamic temperature as

$$\frac{1}{T} = \left.\frac{\partial S}{\partial E}\right|_{\mathcal{X}}. \tag{2.29}$$

Therefore, two systems that are in thermodynamic equilibrium with one another will have the same temperature, recovering the zeroth law of thermodynamics.

Moreover, the entropy variation of the joint system before contact and after reaching thermal equilibrium is given by

$$\Delta S = S_1(\varepsilon) + S_2(\tilde{\varepsilon}) - S_1(E_1) - S_2(E_2) = K_B \ln\left(\frac{\Omega_1(\varepsilon)\Omega_2(\tilde{\varepsilon})}{\Omega_1(E_1)\,\Omega_2(E_2)}\right), \tag{2.30}$$

where the dependence on other state variables were omitted as, by assumption, they do not change during the thermalization process. By construction $\Omega_1(\varepsilon)\Omega_2(\tilde{\varepsilon})$ is the maximum value of the product $\Omega_1\Omega_2$, hence $\Delta S \geq 0$, recovering a consequence of the second law. This is equivalent to saying that the system will evolve to the joint state with the biggest number of accessible joint microstates.

During the thermalization process, the variation in entropy of the joint system between neighboring states is given by

$$\delta S = \left(\left.\frac{\partial S_1}{\partial \varepsilon}\right|_{\mathcal{X}_1} - \left.\frac{\partial S_2}{\partial \tilde{\varepsilon}}\right|_{\mathcal{X}_2}\right)\delta\varepsilon = \left(\frac{1}{T_1} - \frac{1}{T_2}\right)\delta\varepsilon, \tag{2.31}$$

which is zero when the condition (2.28) is satisfied. As the system is evolving to satisfy that condition, the quantity in brackets must approach zero as the system approach thermodynamic equilibrium, then, if $T_1 > T_2$, the temperature T_1 will decrease as the temperature

T_2 increase. On the other hand, if $T_1 < T_2$, the temperature T_1 will increase while T_2 decrease. With that in mind, one should establish the relation between temperature and the energy of the system. Most systems in nature do not have an upper bound for the energy. Then, the more energy one furnishes, the more microstates will be accessible. Therefore, the entropy is a monotonically increasing function of energy, implying that temperature is also a monotonically increasing function of energy. One can then say that energy flows from the system with the higher temperature to the system with the smaller temperature, this statement can be shown to be equivalent to Kelvin's statement of the second law 2.1.1. The discussion on the possibility of negative absolute temperatures will be delayed to the next section.

For the first law, consider a system in the microcanonical ensemble with energy E and a set $\mathcal{X}$ of known values of state variables. Then consider variations of the Boltzmann entropy by reversibly changing $\mathcal{X}$. The amount of work done to change $\mathcal{X}$ is $đW = \mathcal{F} \cdot \delta\mathcal{X}$[7], where $\mathcal{F}$ is the set of values of generalized forces associated with the state variables $\mathcal{X}$. The first-order variation of the entropy is given by

$$\delta S = S\left(E + \mathcal{F} \cdot \delta\mathcal{X}, \mathcal{X} + \delta\mathcal{X}\right) = \left(\left.\frac{\partial S}{\partial E}\right|_{\mathcal{X}} \mathcal{F} + \left.\frac{\partial S}{\partial \mathcal{X}}\right|_{E} \right) \cdot \delta\mathcal{X}. \tag{2.32}$$

In thermodynamic equilibrium, the entropy will not change, therefore, $\delta S = 0$ implying that

$$\left.\frac{\partial S}{\partial \mathcal{X}}\right|_{E} = -\mathcal{F} \left.\frac{\partial S}{\partial E}\right|_{\mathcal{X}} = -\frac{\mathcal{F}}{T}, \tag{2.33}$$

[7]Note that $\mathcal{F} \cdot \delta\mathcal{X}$ is a compact way of writing $\sum_i \mathcal{F}_i \delta\mathcal{X}_i$, where $\mathcal{F}_i$ is i'th generalized force and $\mathcal{X}_i$ is the associated generalized displacement.

then, one can write an infinitesimal entropy variation $\mathrm{d}S$ as

$$\mathrm{d}S = \frac{\mathrm{d}E}{T} - \frac{\mathcal{F} \cdot \mathrm{d}\mathcal{X}}{T}, \tag{2.34}$$

hence

$$\mathrm{d}E = T\,\mathrm{d}S + \mathcal{F} \cdot \mathrm{d}\mathcal{X}\,. \tag{2.35}$$

The heat intake of the system can be identified as $đQ = T\,\mathrm{d}S$. We have then recovered a version of the first law of thermodynamics.

The third law comes from the fact that at absolute zero temperature, the system is in its ground state, which might be degenerated. Then, the entropy of a system at absolute zero temperature is $S = K_B \ln g$, where g is the degeneracy of the system, recovering a version of the third law of thermodynamics.

2.2.3 Canonical ensemble

Let S be a system at temperature T composed of N particles occupying a volume V. The macrostate of the system is $M\,(N, V, T)$ and the set of microstates compatible with $M\,(N, V, T)$ forms the canonical ensemble. It is not possible to treat this system as a microcanonical ensemble without knowing the energy of the system. In order to treat this case it is necessary to consider an auxiliary system, a heat reservoir R at a temperature T and energy $E_r \gg \tilde{E}$, where $\tilde{E}$ is the energy of the system S when it is in thermodynamic equilibrium with R, this situation is illustrated in figure 2.3. The assumption about the energy of the reservoir is necessary to ensure that its temperature will not change by any significant amount after reaching thermal equilibrium with S.

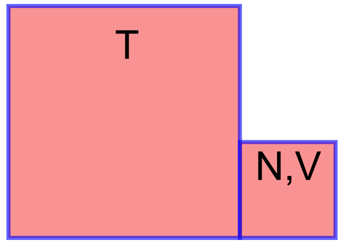

Figure 2.3: Heat reservoir R (left) in thermodynamic equilibrium with the system S. As a consequence of the thermodynamic equilibrium, the system S is also at the temperature T ensuring that it will be in the desired macrostate $M\left(N,V,T\right)$.

As we have seen, partitioning a system in the microcanonical ensemble results in subsystems whose macrostates are the ones with the highest number of accessible microstates, which is also the state with the highest entropy given the available energy of the full system. Hence, the equilibrium state of the canonical ensemble is such that it maximizes the entropy given the available energy $E_r + \tilde{E} \approx E_r$. Without the possibility of directly counting the accessible microstates, one can resort to an alternative entropy, called the Von Neumann entropy, which happens to match the Boltzmann entropy we used in the microcanonical situation. The Von Neumann entropy of a system in a quantum state ρ is defined as[8]

$$S\left[\rho\right] = -K_B \operatorname{Tr} \rho \ln \rho = -K_B \sum_i p_i \ln p_i, \tag{2.36}$$

where p_i are the eigenvalues of the density operator ρ. In this section ρ is to be understood as the density operator representing the canonical ensemble. Then, one should solve the problem of maximizing ρ given that the energy of the system is E and the Hamiltonian of the system S is H. A solution for the general case can be found in appendix C.1. Using equations (C.4), (C.4) and (C.6), the density operator ρ, the partition function Z and the

[8]The factor of K_B was inserted here to make the connection with the thermodynamic entropy clearer. The original definition does not have the factor of K_B.

average energy E are given by

$$\rho = \frac{e^{-\beta H}}{Z} \tag{2.37}$$

$$Z = \mathrm{Tr}\, e^{-\beta H} \tag{2.38}$$

$$E = -\frac{\partial \ln Z}{\partial \beta}, \tag{2.39}$$

where $\beta = (K_B T)^{-1}$. Moreover, according to equation (C.8) the thermodynamic entropy is given by

$$S = K_B \ln Z + \frac{E}{T}, \tag{2.40}$$

motivating the identification of the Helmholtz free energy $F = E - TS$ as

$$F = -K_B T \ln Z. \tag{2.41}$$

One should note that the energy of the most probable configuration was defined as $\tilde{E}$ while the average energy of the system is $E = \langle H \rangle$. They are not generally equivalent. To check this, one can calculate the variance of the energy ΔE as

$$\Delta E = \left\langle H^2 \right\rangle - \langle H \rangle^2 = \frac{\partial^2 \ln Z}{\partial \beta^2} = K_B T^2 \left. \frac{\partial E}{\partial T} \right|_{N,V}. \tag{2.42}$$

The energy of the system is an extensive quantity, $\Delta E \propto N$, and as a consequence, the relative error $\frac{E}{\sqrt{\Delta E}} \propto \sqrt{N}^{-1}$. For the usual systems that are considered, $N \sim 10^{23}$ and the error becomes negligible. In that case, $E = \tilde{E}$ and the canonical ensemble is equivalent to the microcanonical ensemble.

2.2.4 Grand Canonical Ensemble

There are many situations where the number of particles is not conserved, and the energy of the system is not known. In those situations, one should use the grand canonical ensemble. Let S be a system with a known chemical potential μ, temperature T, and volume V. The macrostate of the system is $M(\mu, V, T)$ and the set of compatible microstates forms the grand canonical ensemble.

The approach used to determine the density operator of the grand canonical ensemble is similar to the one used for the canonical ensemble. The only difference is that now the reservoir provides both the temperature and the chemical potential, as presented in figure 2.4. The reservoir R is at a temperature T and chemical potential μ, with energy $E_R \gg \tilde{E}$ and number of particles $N_r \gg \tilde{N}$, where $\tilde{E}$ and $\tilde{N}$ are the thermodynamic equilibrium energy and particle number of the system S. The equilibrium state of the grand canonical ensemble is such that the entropy is maximized given the available energy $E_r + \tilde{E} \approx E_r$ and number of particles $N_r + \tilde{N} \approx N_r$.

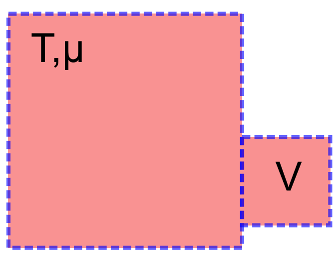

Figure 2.4: Heat and particle reservoir R (left) in thermodynamic equilibrium with the system S (right). S exchanges heat and particles with the R. As a consequence of the thermodynamic equilibrium, the system S is also at the temperature T and chemical potential μ ensuring that it is in the macrostate $M(\mu, V, T)$.

The grand canonical ensemble is equivalent to a situation where the average energy and the average number of particles are given, hence, the Hamiltonian H and the particle number operator $\hat{N}$ are given. Using equations (C.4), (C.4) and (C.6), the density operator ρ, the grand partition function $\mathcal{Q}$, the average number of particles N and the average energy E are given by

$$\rho = \frac{e^{-\beta H+\alpha N}}{\mathcal{Q}} \tag{2.43}$$

$$\mathcal{Q} = \mathrm{Tr}\left(e^{-\beta H+\alpha N}\right), \tag{2.44}$$

$$N = \frac{\partial \ln \mathcal{Q}}{\partial \alpha}, \tag{2.45}$$

$$E = -\frac{\partial \ln \mathcal{Q}}{\partial \beta}, \tag{2.46}$$

where $\beta = (K_B T)^{-1}$ and $\alpha = \mu\beta$. Moreover, according to equation (C.8) the thermodynamic entropy is given by

$$S = K_B \ln Z + \frac{E}{T} - \frac{\mu N}{T}, \tag{2.47}$$

motivating the identification of the grand potential $\Phi = E - TS - \mu N$ as

$$\Phi = -K_B T \ln \mathcal{Q}. \tag{2.48}$$

As in the canonical ensemble, $\tilde{E}$ and $\tilde{N}$ are respectively the energy and particle number of the most probable partition, but they are not necessarily equal to the average energy E and particle number N. The former values can be evaluated. The variance of the average number of particles is

$$\Delta N = \left\langle N^2 \right\rangle - N^2 = \frac{\partial^2 \ln Z}{\partial \alpha^2} = \frac{\partial N}{\partial \alpha}. \tag{2.49}$$

The number of particles is an extensive quantity, therefore, $\Delta N \propto N$ and the relative error $\frac{N}{\sqrt{\Delta N}} \propto \sqrt{N}^{-1}$. Hence, for the usual systems, $N \sim 10^{23}$ and the deviation from $\tilde{N}$ is negligible. The same happens for the deviation from $\tilde{E}$, therefore, all the ensembles are equivalent in this limit.

2.2.5 The issue of "negative absolute temperatures"

Using the laws of equilibrium thermodynamics as a basis may lead to the conclusion that negative absolute temperatures are impossible because the heat transfer would violate the second law of thermodynamics 2.1.1 and the third law of thermodynamics 2.1.

Naively, one might think that negative temperatures violate the third law because to reach a negative temperature it seems to be necessary to pass through the absolute zero, which according to the third law of thermodynamics is unattainable. That is not necessarily true. According to statistical mechanics, the temperature can also be defined as

$$\frac{1}{T} = \left.\frac{\partial S}{\partial E}\right|_{\mathcal{X}}, \tag{2.50}$$

hence, if the entropy decreases when the system absorbs energy, the temperature will be negative. Such behavior is known to exist in systems that have an upper bound for the energy, an example is spin systems, where this phenomenon has been measured for the first time, as reported in [36]. More careful reasoning will indicate that there is no need to go through absolute zero in order to reach negative temperatures. To understand, one should think of the behavior of $\frac{\partial S}{\partial E}$. When the entropy is near its global maximum, adding more energy will cause a small effect. Therefore, the derivative is close to zero, indicating a very high temperature. When the entropy reaches its maximum, the derivative is zero, and the temperature tends to infinity. After that point, the entropy decreases when more energy is added to the system, and the derivative will be a small negative number. Hence, the temperature suddenly jumps from $+\infty$ to $-\infty$, not passing by zero. Therefore, no violation of the third law is necessary.

By accepting negative temperatures, one now faces a problem with the second law of thermodynamics 2.1. By putting a negative temperature system at temperature T_1 in contact with a positive temperature system at temperature T_2, the heat will flow from the negative temperature system into the positive temperature system as the former has more energy than the second one. Hence, in this unusual situation, the heat flow from the lower temperature system into the higher temperature system, and the second law of thermodynamics is violated. In those situations, one should use a modified version of Kelvin's statement 2.1.1 of the second law of thermodynamics that also takes this particular problem into account [37]. On the other hand, if both systems were at a negative temperature, the usual statement of the second law still holds.

Hence, there is no drama with the laws of thermodynamics and negative temperatures. By the time the previously cited experiment was done, discussion on the plausibility of negative absolute temperatures happened, resulting in an article by N. F. Ramsey, where he explains how thermodynamics and statistics behave at negative absolute temperatures [37]. Even after that article, there was some resistance which resulted in another article by M. J. Klein [38] defending Ramsey's point of view. The issue seemed solved until 2013 when a modern experiment that measured negative absolute temperature was released [39]. This article revived the discussion as can be seen in the article [40] where the authors argued that negative temperatures were an artifact of Boltzmann's definition of entropy and by using

Gibbs's definition of entropy [33] the negative temperatures would not be predicted. There was an intense debate on Gibbs's vs. Boltzmann entropy definition, but in 2018 the debate seems to have come to an end with an article by R. Swendsen [41] where, among other issues, the Boltzmann definition of entropy was shown to be more appropriate than Gibbs's.

2.2.6 "Violations" of the laws of thermodynamics

After presenting the statistical mechanical approach, there are some points to highlight, concerning the laws of thermodynamics. While the beauty of statistical mechanics lies in how it makes the laws of thermodynamics emerge from microscopic physics, it also comes at a price, but a very worthy one. Strictly considering statistical mechanics, the laws of thermodynamics arise as statistical laws and are valid under a specific set of assumptions such as the postulate of equal a priori probabilities and relatively weak short-range interactions. Therefore those laws are vulnerable to violations due to their statistical nature and prone to extensions that may try to relax the assumptions of short-range and relatively weak interactions.

Different authors treat the basic notion of thermodynamic equilibrium with reservations. One example is J.Beattie and I. Oppenheim, who say in their book [42] that:

> Insistence on a strict interpretation of the definition of equilibrium would rule out the application of thermodynamics to practically all states of real systems.

To understand the motivation for that thought, we can quote another author, A.B Pippard, who gives a remarkable example [43]:

> Given long enough a supercooled vapour will eventually condense of its own accord, and given long enough a mixture of hydrogen and oxygen will transform itself into water. The time involved may be so enormous, however, perhaps 10^{100} years or more, that the process is not perceptible. For most purposes, provided the rapid change is not artificially stimulated, the systems may be regarded as being in equilibrium.

Therefore, according to the definition of thermodynamic equilibrium 2.1.1, the systems in the example will never reach thermodynamic equilibrium if one is considering observing the system for 10^{100} years or more. On that time scale, one will witness the spontaneous appearance of net macroscopic flows of energy and matter within the system. Hence, one should be aware that there is a limitation in the notion of thermodynamic equilibrium, related to the fact that it is expected to persist for a long but finite time.

To explain the spontaneous break of thermodynamic equilibrium, we can re-visit the first argument presented in section 2.2.2. What was determined was that there is not the equilibrium partition of energy between subsystems but rather the most probable one. As a consequence of the postulate of equal a priori equilibrium probabilities, every partition of energy between the subsystems is equally likely to happen. The main point of the argument is that the number of microstates associated with the most probable partition of energy is much bigger than the number of microstates associated with any other partition. Hence, the system will spend much more time in the most probable partition. Still, it is possible for the system to evolve to another partition but it is very improbable. Notably, if the system evolves to any partition of energy other than the most probable, and stays in that configuration for long enough such that it can be considered in thermodynamic equilibrium, its entropy would have suffered a reduction, violating the second law of thermodynamics.

According to statistical mechanics, the entropy itself is indeed allowed to diminish. What is true, however, is that the ensemble-averaged entropy will always increase or stay constant. Nevertheless, even this more careful statement is not violation-proof. This reservation is based on the relatively new Fluctuation Theorems(FT), whose first version came out in 1993 in [44], being improved in [45]. For a pedagogical presentation of FT, the reader is referred to [46]. The FT provides an estimation for the probability of a system observed during a time Δt to violate the second law of thermodynamics. Fortunately, such probability becomes exponentially smaller as Δt or the system size becomes larger, recovering the expected behavior of the ensemble-averaged entropy. An experiment [47] was conducted in 2002 confirming the FT, and the group also found that for Δt larger than a few seconds, no violations can be observed, as predicted from the second law of thermodynamics. For references on other experiments testing the FT, one can refer to [48].

Beyond the "violations" of the second law presented above, in situations where the number of particles is small enough to make the relative uncertainty in the steepest descent approximation considerable, violations of the zeroth law will happen as a consequence of fluctuations from the most probable state. As a consequence, the transitivity of thermodynamic equilibrium is not always true. The small number of particles in a system is not the only cause of such violationos, as presented in [49], [50], [51], and [52].

Finally, one should not worry about violations of the laws of thermodynamics if the system in question has a significant number of particles and interact with the bath by relatively weak short-range interactions.

2.3 Maxwell's demon, Quantum Information and Thermodynamics

In section 1.6 we presented the concept of information as developed by Shannon in [18]. Up to now, the connection between Information and Physics is anchored in the similarity between Shannon entropy 1.40 and Von Neumann entropy D.3. But there is a deeper relation between Information and Physics. To present that link to the reader we will use the famous Maxwell's Demon problem.

To recover the historical aspect of the original problem, we will use an adapatation of the original statement. The original statement was presented in a letter Maxwell sent to Tait in 1867 which was transcribed in a book about Tait's life [53]. Our adaptation[9] of the original statement is as follows:

> Let A and B be two vessels divided by a diaphragm and let them contain elastic molecules in a state of agitation which strike each other and the sides. Let the number of particles be equal in A and B but let those in A have the greatest energy of motion.
>
> Now conceive a finite being who knows the paths and velocities of all the molecules by simple inspection but who can do no work except open and close a hole in the diaphragm by means of a slide without mass.
>
> Let him first observe the molecules in A and when he sees one coming the square of whose velocity is less than the mean sq. vel. of the molecules in B let him open the hole and let it go into B. Next let him watch for a molecule of B, the square of whose velocity is greater than the mean sq. vel. in A, and

[9]We have only removed a few words of the original text.

when it comes to the hole let him draw the slide and let it go into A, keeping the slide shut for all other molecules.

Then the number of molecules in A and B are the same as at first, but the energy in A is increased and that in B diminished, that is, the hot system has got hotter and the cold colder and yet no work has been done, only the intelligence of a very observant and neat-fingered being has been employed.

Or in short if heat is the motion of finite portions of matter and if we can apply tools to such portions of matter so as to deal with them separately, then we can take advantage of the different motion of different proportions to restore a uniformly hot system to unequal temperatures or to motions of large masses.

Only we can't, not being clever enough

To model the situation proposed by Maxwell we can consider that the vessels A and B, both with volume V[10], are each filled with an ideal gas composed of N particles of mass m. The net energy of the particles in the vessel A is E_A and that of the particles in the vessel B is E_B, additionally, let $E_A > E_B$ and $E = E_A + E_B$.

The entropy of an ideal gas of N free particles of mass m and net energy E enclosed in a vessel of volume V is given by [27]

$$S\left(E,V,N\right) = NK_B\left[\frac{5}{2} + \frac{3}{2}\ln\left(\frac{4\pi m}{3}\right) + \ln\left(\frac{V}{N}\right) + \frac{3}{2}\ln\left(\frac{E}{N}\right)\right]. \qquad (2.51)$$

As the action of the Demon can only change the energy, we conveniently write this entropy

[10] Equal volumes for simplicity.

as

$$S(E) = C(N,V) + NK_B \frac{3}{2} \ln\left(\frac{E}{N}\right), \quad \text{where} \tag{2.52}$$

$$C(N,V) = NK_B \left[\frac{5}{2} + \frac{3}{2} \ln\left(\frac{4\pi m}{3}\right) + \ln\left(\frac{V}{N}\right)\right] \tag{2.53}$$

The term $C(N,V)$ is kept constant by actions of the Demon. Additionally, the temperature of an ideal gas of N particles with net energy E is [27]

$$T(E,N) = \frac{2}{3} \frac{E}{NK_B} \tag{2.54}$$

The entropies in Maxwell's initial setup are $S_A(E_A)$ and $S_B(E_B)$ with $E_A > E_B$, which implies $T_A > T_B$ as can be noted by looking at equation (2.54). Due to the actions of the Demon over an interval of time, the net energy of system A increases to E'_A and that of B decreases to E'_B, leading to an increase in the temperature of system A and a decrease in the temperature of system B, therefore violating the second law of thermodynamics as claimed by Maxwell.

We now depart from Maxwell's point, that was more qualitative and focused on the temperature, for a more quantitative and entropy-oriented approach. The net entropy of the system in the initial setup is

$$S_{AB}(E_A, E_B) = S_A + S_B = C(2N, 2V) + \frac{3}{2} NK_B \ln\left(\frac{E_A E_B}{N^2}\right). \tag{2.55}$$

If there was no Demon and the hole were left open, the system would flow towards its thermodynamic equilibrium state and in that situation the energy would be homogeneously distributed in such a way that $E_A = E_B = \dfrac{E}{2}$ for the given situation. The thermodynamic entropy in that situation is given by

$$S_{AB}\left(\frac{E}{2},\frac{E}{2}\right) = C\left(2N,2V\right) + 3NK_B \ln\left(\frac{E}{2N}\right). \tag{2.56}$$

Then, the difference between the thermodynamic equilibrium entropy and the initial setup entropy is

$$\begin{aligned} S_{AB}\left(\frac{E}{2},\frac{E}{2}\right) - S_{AB}\left(E_A,E_B\right) &= 3NK_B \ln\left(\frac{E}{2N}\right) - \frac{3}{2}NK_B \ln\left(\frac{E_A E_B}{N^2}\right) \\ &= 3NK_B \ln\left(\frac{E}{2\sqrt{E_A\left(E-E_A\right)}}\right). \end{aligned} \tag{2.57}$$

That difference is minimized to zero when $E_A = \dfrac{E}{2}$, therefore, the thermodynamic equilibrium situation has a bigger entropy for any $E_A < E$, as expected. Also, by following the reasoning presented in the last paragraph, it is easy to conclude that the natural flow of energy would be from system A to system B. We now proceed to calculate by how much the Demon decreases the net entropy of the system after one complete operation.

Consider that the Demon operated the hole, letting one particle with velocity $\vec{v}_1$, such that $|\vec{v}_1|^2 > \left\langle|\vec{v}_A|^2\right\rangle$, escape from system B into system A and another particle with velocity $|\vec{v}_2|^2$, such that $|\vec{v}_2|^2 < \left\langle|\vec{v}_B|^2\right\rangle$, escape from system A into system B. Given that $E = N\left\langle|\vec{v}|^2\right\rangle$, it is easy to see that $E_A > E_B$ implies $|\vec{v}_1|^2 > |\vec{v}_2|^2$. By acting that way, the

Demon changes the energies of the systems A and B as follows:

$$E_A \rightarrow E'_A = E_A + \frac{m}{2}\left(\vec{v}_1^2 - \vec{v}_2^2\right) = E_A + \epsilon \quad \text{and} \tag{2.58}$$

$$E_B \rightarrow E'_B = E_B - \frac{m}{2}\left(\vec{v}_1^2 - \vec{v}_2^2\right) = E_B - \epsilon. \tag{2.59}$$

The variation in the entropy of S_A produced by that increase in E_A is

$$\Delta S_A = S\left(E_A + \epsilon\right) - S\left(E_A\right) = NK_B \frac{3}{2} \ln\left(1 + \frac{\epsilon}{E_A}\right) > 0, \tag{2.60}$$

and the variation in the entropy of S_B produced by the decrease in E_B is

$$\Delta S_B = S\left(E_B - \epsilon\right) - S\left(E_A\right) = NK_B \frac{3}{2} \ln\left(1 - \frac{\epsilon}{E_B}\right) < 0. \tag{2.61}$$

As $E_A > E_B$, it is easy to see that $|\Delta S_A| > |\Delta S_B|$. Then, the variation in the net entropy of the system is

$$\begin{aligned}\Delta S_{AB} &= S_{AB}\left(E_A + \epsilon, E_B + \epsilon\right) - S_{AB}\left(E_A, E_B\right) = \Delta S_A + \Delta S_B \\ &= NK_B \frac{3}{2} \ln\left(1 + \frac{\epsilon}{E_A} - \frac{\epsilon}{E_B} - \frac{\epsilon^2}{E_A E_B}\right) < 0,\end{aligned} \tag{2.62}$$

as $\frac{\epsilon}{E_A} < \frac{\epsilon}{E_B}$, supporting the previous claim that the net entropy of the system decreases.

The caveat to the argument above is that the Demon's knowledge was neglected. That is, in order to act as Maxwell proposed, the Demon must keep track of the velocity of molecules in each side, in other words, a blind demon might not be able to produce the entropy reduction effect. Therefore, to correctly handle the net entropy variation of the system, one would have to include both: the variation of entropy in the Demon's memory and the variation of the entropy that is caused by whatever measuring apparatus the demon is using. A precise account for those factors produced a very rich discussion about that specific topic. The interested reader is referred to the book [54] where an extensive treatment of several discussions regarding solutions to Maxwell's Demon are presented.

By using Maxwell's Demon we provided arguments to motivate the connection between information and physics, that is, Shannon's information theory can be used as an additional framework to gain insight about physical phenomena. Despite that reasoning, it is not true that the relation between information and physics is a consensus among physicists. The interested reader is referred to [55, 56, 57, 58].

3

General Relativity and Black hole Thermodynamics

When one refers to the laws of Black hole thermodynamics, one is referring to the laws of Black hole mechanics which are formally analogous to the laws of thermodynamics. The fact that, in the classical level, the Black hole area does not decrease, as proven by Hawking in [59], together with the first law of Black hole mechanics, suggests that there might be a relation between the Black hole area and thermodynamic entropy. In this chapter, the laws of Black hole mechanics will be presented, together with a brief introduction to general relativity. This chapter is organized as follows: In section 3.1, a brief review of general relativity is presented. In section 3.2 we review the formalism of Penrose diagrams. In section 3.3 definitions regarding the causal structure of spacetime are presented in order to properly state the information loss puzzle in the next chapter. In section 3.4 we present a rationale that allows one to have a notion of time and space within the framework of general relativity. The notion of time plays a significant role when it comes to quantizing fields on curved spacetimes. In section 3.5, the laws of Black hole thermodynamics are presented. In section 3.6, the reader is presented to a schematic approach of the semi-classical framework.

3.1 Mathematical structures of General Relativity

In this section, a minimal presentation of the mathematical structures directly connected to the formalism of general relativity will be shown. While several elements from differential geometry will be used, a complete introduction of those concepts would fall out of the scope of this text. For a complete text on differential geometry, the reader is referred to [60] and [61]. For a complete text on general relativity, the reader is referred to [62], [63].

The primordial concept of general relativity is that of spacetime, which is a generalization of Euclidean space, i.e., the usual 3-dimensional space. A spacetime is defined as a pair $(\mathcal{M}, \mathbf{g})$ where $\mathcal{M}$ is a n-dimensional smooth manifold with $n \geq 2$ and $\mathbf{g}$ is a metric on

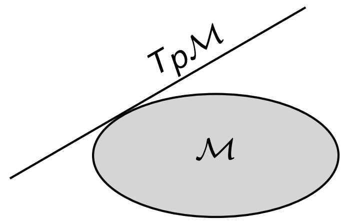

Figure 3.1: A manifold $\mathcal{M}$ and the tangent space $T_p\mathcal{M}$ at a point $p \in \mathcal{M}$.

$\mathcal{M}$. The manifold plays the role of a region of the universe while the metric has information on how to measure distances on such manifold. We will work exclusively with pseudo-Riemannian metrics, which are defined as:

Definition 3.1.1 (Pseudo-Riemannian metric). (Adapted from [60]) A pseudo-Riemannian metric $\mathbf{g}$ on $\mathcal{M}$ is a type(0,2) tensor field on $\mathcal{M}$ which satisfies the following axioms at each point $p \in \mathcal{M}$:

1. $\mathbf{g}_p\left(X, Y\right) = \mathbf{g}_p\left(Y, X\right)$
2. if $\mathbf{g}_p\left(X, Y\right) = 0$ for any $X \in T_p\mathcal{M}$, then $Y = 0$,

where $T_p\mathcal{M}$ is a tangent space to the manifold $\mathcal{M}$ at a point $p \in \mathcal{M}$. Here $X, Y \in T_p\mathcal{M}$ and $\mathbf{g}_p = \mathbf{g}\Big|_p$. In short, $\mathbf{g}_p$ is a symmetric positive-definite bilinear form.

A visual representation of the idea of tangent space is presented in figure 3.1. Vectors are defined over these tangent spaces. Given $p, q \in \mathcal{M}$, it is not always true that $T_q\mathcal{M} = T_p\mathcal{M}$. Therefore, it is not trivial to relate vectors at different points of a manifold. On flat spacetimes it is trivial because they are such that the tangent space is the same everywhere, that is, given a flat spacetime $(\mathcal{A}, \mathbf{g})$, then

$$T_p\mathcal{A} = T_q\mathcal{A} \;\forall\; q, p \in \mathcal{A}, \tag{3.1}$$

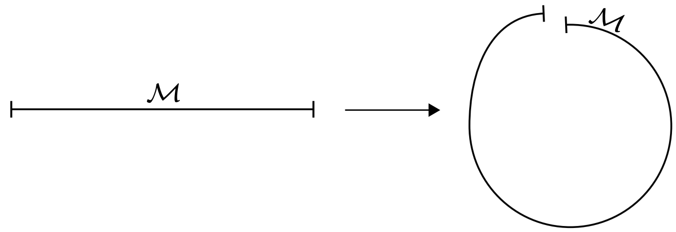

Figure 3.2: A continuous deformation of a straight line manifold $\mathcal{M}$ into a circular shape.

and because of that one need not worry when dealing with vector operations on flat spacetimes. Hence, the necessity of the more sophisticated formalism is related to the fact that most spacetimes in general relativity are not flat, they are curved spacetimes.

It is not trivial to tell if two spacetimes are different because any continuous deformation of a spacetime produces another spacetime with the same properties as the former one, therefore, it produces an equivalent spacetime.One can understand a continuous deformation as a transformation that does not tear, stick, or fold the spacetime. For instance, one can deform a straight line into a circular shape, as shown in figure 3.2. However, following the formalism of differential geometry, the circular shape is as flat as the straight line. The appropriate way of distinguishing between truly curved spacetimes and disguised flat spacetimes is by calculating the Riemann tensor, defined as

$$R : \Xi\left(\mathcal{M}\right) \otimes \Xi\left(\mathcal{M}\right) \otimes \Xi\left(\mathcal{M}\right) \mapsto \Xi\left(\mathcal{M}\right) \tag{3.2}$$

$$R\left(X, Y, Z\right) \equiv \boldsymbol{\nabla}_X \boldsymbol{\nabla}_Y Z - \boldsymbol{\nabla}_Y \boldsymbol{\nabla}_X Z - \boldsymbol{\nabla}_{[X,Y]} Z, \tag{3.3}$$

where $\Xi\left(\mathcal{M}\right)$ is the set of vector fields on $\mathcal{M}$, and $\boldsymbol{\nabla} : \Xi\left(\mathcal{M}\right) \times \Xi\left(\mathcal{M}\right) \mapsto \Xi\left(\mathcal{M}\right)$ are the Levi-Civita connections. The spacetime is said to be flat if and only if the Riemann tensor vanishes identically.

Using the Riemann tensor, one can define two additional quantities, the Ricci tensor[60],

$$Ric : \Xi(\mathcal{M}) \otimes \Xi(\mathcal{M}) \mapsto \mathbb{R} \tag{3.4}$$

$$Ric(X,Y) \equiv \langle \mathrm{d}x^{\mu}, R(e_{\mu}, Y, X) \rangle , \tag{3.5}$$

and the Ricci Scalar[60]

$$\mathcal{R} : \Xi(\mathcal{M}) \otimes \Xi(\mathcal{M}) \mapsto \mathbb{R} \tag{3.6}$$

$$\mathcal{R}(X,Y) \equiv g^{\mu\nu} Ric(e_{\mu}, e_{\nu}) . \tag{3.7}$$

These two quantities are then used to define the Einstein tensor

$$G(X,Y) = Ric(X,Y) - \frac{\mathbf{g}(X,Y)}{2} \mathcal{R}(X,Y) . \tag{3.8}$$

The Einstein tensor is the mathematical quantity used to relate matter and geometry. This is done by the so-called Einstein Equation

$$G(X,Y) = \frac{8\pi G}{c^4} T(X,Y) , \tag{3.9}$$

where G is Newton's gravitational constant, c is the speed of light and T is the energy-momentum tensor of matter.

Conformal diagrams are a way of representing the spacetime which makes it much easier to understand its causal structure. When analyzing the information loss puzzle, we will make use of them. Moreover, conformal diagrams offer a finite representation of spacetime. These diagrams are also called Penrose or Carter-Penrose diagrams as Penrose introduced the idea in [64] and Carter in [65].

The idea of a conformal diagram is the following: For a given physical metric **g** and a basis $\{e_\mu\}$ for $\Xi(\mathcal{M})$, the displacement[1] is $\mathrm{d}s^2 = \mathbf{g}(e_\mu, e_\nu) = g_{\mu\nu}\,\mathrm{d}x^\mu\,\mathrm{d}x^\nu$. Then, the objective is to find a coordinate system where that displacement can be written as $\mathrm{d}s^2 = \Omega^2\,\mathrm{d}\tilde{s}^2$, such that $\Omega = 0$ on asymptotic regions of $\mathcal{M}$. We call $\mathrm{d}\tilde{s}$ the conformal form of that displacement. To find it, one usually goes to null (or light-cone) coordinates, then compactify them using a smooth function $f : (-\infty, \infty) \mapsto (a, b)$, where a, b are finite. Two commonly used functions are tan and $tanh^{-1}$. Then one makes a suitable transformation in order to achieve the form $\mathrm{d}s^2 = \Omega^2\,\mathrm{d}\tilde{s}^2$. The full details about the process of evaluating Ω are involved and fall out of the scope of this text. For a complete treatment of this topic the reader is referred to [66] and [67].

The conformal form of the metric is useful because the light-cones are preserved, regardless of how complicated the expression of Ω is. Hence, a conformal diagram is a tool to study the causal structure of spacetime. As an example, we can look at Minkowski space-

[1] In the literature it is common to call it the metric, even thought it is an abuse of language.

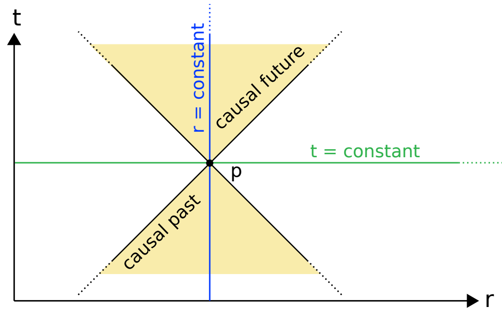

Figure 3.3: A drawing of a region of spacetime near an event p. The regions in yellow are the interior of a light-cone centered on p. In blue is a hypersurface of constant r and in green is a hypersurface of constant t.

time, whose metric in spherical coordinates is given by

$$\begin{cases} \mathrm{d}s^2 = -c^2\,\mathrm{d}t^2 + \mathrm{d}r^2 + r^2\left(\mathrm{d}\theta^2 + \sin\left(\theta\right)^2\mathrm{d}\phi^2\right), \\ t \in \mathbb{R},\; r \in \mathbb{R}_{\geq 0},\; \theta \in [0, \pi[\;, \phi \in [0, 2\pi[\,. \end{cases} \tag{3.10}$$

For simplicity, we take θ and ϕ as constants and look only at a 2-dimensional surface of spacetime. There is no loss of generality due to the spherical symmetry of Minkowski spacetime. The metric on a 2-dimensional hypersurface of constant θ and ϕ is

$$\mathrm{d}s^2 = -c^2\,\mathrm{d}t^2 + \mathrm{d}r^2\,. \tag{3.11}$$

In this coordinate system a drawing of the full space-time would demand an infinite sheet of paper as presented in figure 3.3. By changing to new coordinates $\psi \in (-\pi, \pi)$ and $\xi \in$

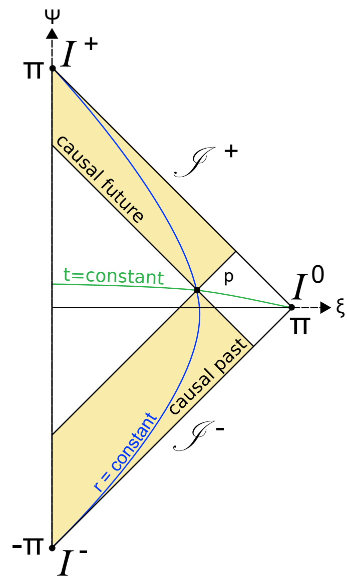

Figure 3.4: The conformal diagram for Minkowski spacetime. The hyper-surfaces of constant t and r were deformed due to the change in coordinates but not the light-cone.

$[0, \pi)$, defined as

$$\begin{cases} t + r = \tan\left(\dfrac{\psi + \xi}{2}\right) \\ t - r = \tan\left(\dfrac{\psi - \xi}{2}\right) \end{cases}, \tag{3.12}$$

the metric is cast into the form $\mathrm{d}s^2 = \Omega^2\left(-\mathrm{d}\psi^2 + \mathrm{d}\xi^2\right)$, where

$$\Omega = \frac{1}{2}\sec\left(\frac{\psi + \xi}{2}\right)\sec\left(\frac{\psi - \xi}{2}\right). \tag{3.13}$$

The conformal metric is then $\mathrm{d}\tilde{s} = -\mathrm{d}\psi^2 + \mathrm{d}\xi^2$. As both coordinates are bounded in a finite interval, it is possible to draw the whole spacetime in a finite picture, the so-called conformal(or Penrose) diagram, presented in figure 3.4. The advantage is that it is now possible to represent infinity while keeping the light-cone structure, making it simpler to study the causal structure of a spacetime. The infinities represented in a conformal diagram are defined as [62]:

- I^+ is the future timelike infinity, the region $t \to \infty$ at finite radius r (region toward which timelike geodesics extend).
- $\mathscr{I}^+$ is the future null infinity, the region $t + r \to \infty$ at finite time $t - r$ (region toward which outgoing null geodesics extend).
- I^0 is the spacelike infinity, the region $r \to \infty$ at finite time t (region toward which spacelike slices extend).
- $\mathscr{I}^-$ is the past null infinity, the region $t - r \to -\infty$ at finite $t + r$ (region from which ingoing null geodesics come).
- I^- past timelike infinity, the region $t \to -\infty$ at finite radius r (region from which timelike geodesics come).

The behavior of these infinities is valid in any spacetime where they do exist. Because of that, the conformal diagram of any asymptotically flat spacetime is expected to resemble that of Minkowski spacetime.

Another interesting example is Schwarzschild spacetime for which the metric is given by

$$\begin{cases} \mathrm{d}\tilde{s}^2 = -c^2\left(1 - \frac{r_s}{r}\right)\mathrm{d}t^2 + \left(1 - \frac{r_s}{r}\right)^{-1}\mathrm{d}r^2 + r^2\left(\mathrm{d}\theta^2 + \sin\left(\theta\right)^2\mathrm{d}\phi^2\right), \\ t \in \mathbb{R},\ r \in \mathbb{R}_{>0},\ \theta \in [0, \pi]\ , \phi \in [0, 2\pi[\,, \end{cases} \tag{3.14}$$

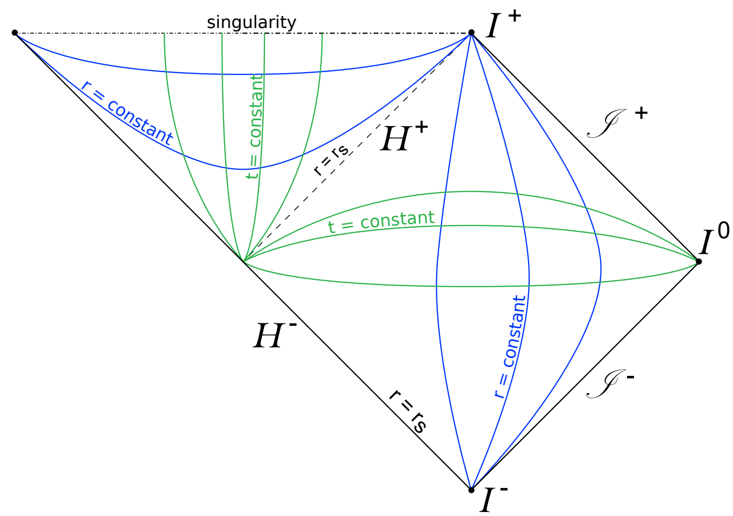

Figure 3.5: The conformal diagram for Schwarzschild spacetime. Note that for $r < r_s$ the constant r and t hyper-surfaces swap roles.

where $r_s = \dfrac{2GM}{c^2}$ is the Schwarzschild radius. The conformal diagram for Schwarzschild spacetime is shown in figure 3.5. For $r < r_s$ the signals of the coefficients of $\mathrm{d}t^2$ and $\mathrm{d}r^2$ flip. Because of that, the otherwise spatial coordinate become a time coordinate and vice-versa. As a consequence, the constant t and r hyper-surfaces swap roles for $r < r_s$. There are also two new special regions:

- H^+ is the future event horizon where $r = r_s$ and $t \to \infty$
- H^- is the past event horizon where $r = r_s$ and $t \to -\infty$

Inside the future event horizon lies the singularity, at $r = 0$. The Black hole region will be precisely defined in the next section. Schwarzschild spacetime has an eternal Black hole that was not formed by the collapse of matter. While useful for studying properties of Black holes, it is not a realistic model of a Black hole formed by the collapse of matter.

Simple reasoning leads to the conformal diagram of a spacetime containing a Black hole

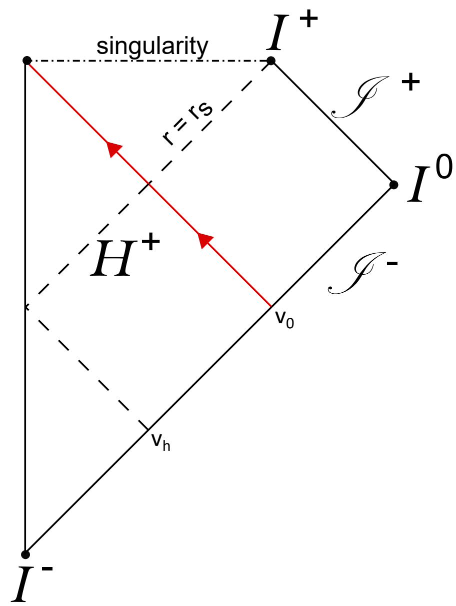

Figure 3.6: The conformal diagram for a spacetime representing a spherical collapse of null matter. The "stitching point" is represented by the red line where otherwise was the future null-infinity.

formed by the collapse of null matter[2]. Given that the collapsing matter is a thin, diffuse, spherically symmetric, neutral and radially collapsing shell of radiation traveling at the speed of light, there is no reason to expect something other than the usual Minkowski spacetime, both inside and outside the shell, before the Black hole forms. After the formation of the Black hole, if there is no more matter outside, there is also no reason to expect anything different from a Schwarzschild Black hole given that there was no charge or angular momentum in the infalling matter. More precisely, according to Birkhoff's theorem, any spherically symmetric vacuum solution of Einstein's equation is a Schwarzschild spacetime [68, 69]. Then, the conformal diagram of such a situation is given by stitching Minkowski and Schwarzschild conformal diagrams over the null curve representing the collapse of the shell of radiation, as presented in figure 3.6. There is no past event horizon as $t \to -\infty$ is

[2]Null matter is any kind of matter that travels at the speed of light, i.e., along null curves.

a region of Minkowski space. That spacetime is an example of a Vaidya spacetime[70]. The general metric for that spacetime is given by

$$\begin{cases} \mathrm{d}\tilde{s}^2 = -c^2\left(1 - \dfrac{2GM\left(v\right)}{c^2 r}\right)\mathrm{d}v^2 + 2\,\mathrm{d}v\,\mathrm{d}r + r^2\left(\mathrm{d}\theta^2 + \sin\left(\theta\right)^2\mathrm{d}\phi^2\right), \\ v \in \mathbb{R},\ r \in \mathbb{R}_{>0},\ \theta \in \left[0, \pi\right],\phi \in \left[0, 2\pi\right[, \end{cases} \tag{3.15}$$

where $M\left(v\right) = 4\pi r^2 \displaystyle\int \mathrm{d}v\, T_{vv}$. It is important to remark that in the general case, v cannot be identified with Schwarzschild nor Minkowski time due to the cross term in the metric. The Vaidya spacetime offers a simple picture where the Hawking effect happens.

3.3 Causal structure of space time

Given a spacetime, one can define several regions that are necessary when describing the information loss puzzle. Furthermore, the global definition of Black hole relies on the causal structure of spacetime. To study those structures, we follow the definitions from [71] and [72]. We will present adapted versions of the most important definitions necessary to understand the information loss puzzle that will be presented in the next chapter. The first set of definitions regards the classification of a vector on the spacetime:

Definition 3.3.1 (Vector types). A vector field $X \in \Xi\left(\mathcal{M}\right)$ at a point $p \in \mathcal{M}$ is

$$\text{spacelike, if } \mathbf{g}_p\left(X, X\right) > 0, \tag{3.16}$$

$$\text{null, if } \mathbf{g}_p\left(X, X\right) = 0, \tag{3.17}$$

$$\text{timelike, if } \mathbf{g}_p\left(X, X\right) < 0. \tag{3.18}$$

Now let $X, Y \in \Xi(\mathcal{M})$ be two non-spacelike vectors at some point $p \in \mathcal{M}$. If we assume that X is future-directed, then Y is

$$\text{future-directed, if } \mathbf{g}_p(X, Y) < 0, \tag{3.19}$$

$$\text{past-directed, if } \mathbf{g}_p(X, Y) > 0. \tag{3.20}$$

Additionally, if $\mathbf{g}_p(X, Y) = 0$, X and Y are null and colinear.

A curve on $\mathcal{M}$ is classified according to its tangent vector.

The second set of definitions formalize the idea of light-cone:

- The chronological future of $\mathcal{S} \subset \mathcal{M}$, denoted $I^+(\mathcal{S})$, is defined as the set of all points $x \in \mathcal{M}$ that can be reached from another point $y \in \mathcal{S}$ by a future-directed timelike curve.

- The causal future of $\mathcal{S} \subset \mathcal{M}$, denoted $J^+(\mathcal{S})$, is defined as the union of $\mathcal{S}$ with the set of all points $x \in \mathcal{M}$ that can be reached from another point $y \in \mathcal{S}$ by a future-directed causal curve[3].

- The chronological past of $\mathcal{S} \subset \mathcal{M}$, denoted $I^-(\mathcal{S})$, is defined as the set of all points $x \in \mathcal{M}$ that can be reached from another point $y \in \mathcal{S}$ by a past-directed timelike curve.

- The causal past of $\mathcal{S} \subset \mathcal{M}$, denoted $J^-(\mathcal{S})$, is defined as the union of $\mathcal{S}$ with the set of points $x \in \mathcal{M}$ that are connected to another point $y \in \mathcal{S}$ by a past-directed causal curve.

[3]A causal curve is a curve that is either timelike or null.

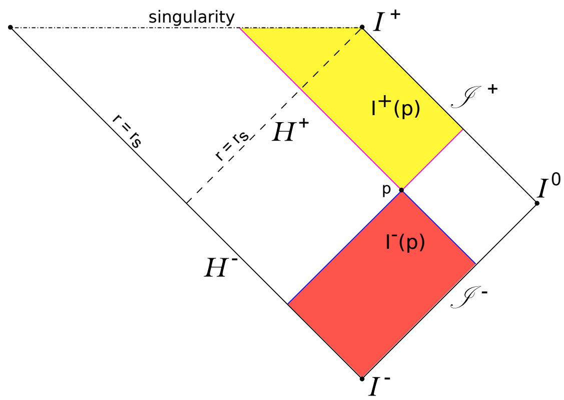

Figure 3.7: In yellow is $I^+(p)$, in red $I^-(p)$. For the causal future, one should take the yellow area and its red boundary. For the causal past, one should take the red area and its blue boundary.

The light-cone with origin at a point $p \in \mathcal{M}$ is then defined as $J^+(p) \cup J^-(p)$. The chronological future (past) of a spacetime region can also be understood as the interior of its future (past) light-cone. And the causal future (past) of that region is its whole (inside and borders) future (past) light-cone. An example is provided in figure 3.7. As the conformal diagrams preserve the light-cones one can use them to reason about these structures on these diagrams.

Given these definitions, one defines the Black hole region $\mathcal{B}$ of an asymptotically flat spacetime $(\mathcal{M}, \mathbf{g})$ as[64]:

$$\mathcal{B} \equiv \mathcal{M} - I^-\left(\mathscr{I}^+\right). \tag{3.21}$$

The boundary of the Black hole region, $\mathcal{H} = \partial\mathcal{B}$, forms the future event horizon, denoted by H^+ in figure 3.7. Similarly, it is possible to define a White hole region and its boundary would be the past event horizon H^- but since that concept will not be used, it will not be presented in this text.

Concerning Black holes, one should be aware that there are several different definitions for it. The presented definition is standard in the context of general relativity, and other areas use different definitions due to the limitations of this one. For instance, this definition requires one to know the behavior of the universe in the future in order to tell if what we have in the center of our galaxy is really a Black hole. Hence, other areas like astrophysics use other definitions of Black holes. A great mathematical reference about the several definitions of Black hole is [73] and a philosophical reference is [74].

Another set of definitions that plays a major role in the information loss puzzle is the one concerning the domains of dependence of regions of spacetime. Those definitions are as follows:

Definition 3.3.2 (Domains of dependence). (Adapted from [72] and [71]) Let S be a set of chronologically disconnected points. Define the future and past domains of dependence of S and the total domain of dependence of S, respectively, as follows:

- $D^{+}(S) = \{x \in \mathcal{M} \mid$ every endless past-directed causal curve containing x meets $S\}$.
- $D^{-}(S) = \{x \in \mathcal{M} \mid$ every endless future-directed causal curve containing x meets $S\}$.
- $D(S) = \{x \in \mathcal{M} \mid$ every endless causal curve containing x meets $S\}$.

It is important to remark that Cauchy development is also known in the literature as domain of dependence. The importance of the domains of dependence is the following: Under the assumption that the underlying physical theory is causal, given the information on a set S of causally disconnected points, one can expect to be able to predict anything on $D^{+}(S)$ and retrodict anything on $D^{-}(S)$. In other words, anything outside the domain of dependence of S cannot be completely predicted using only information on that set. This property derives from the study of the Initial value problem in general relativity.

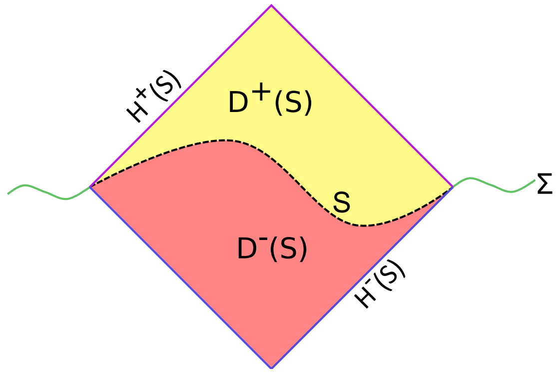

Figure 3.8: The future and past Cauchy developments, $D^{+}(S)$ and $D^{-}(S)$ of a hypersurface $S \subset \Sigma$. Anything outside the Cauchy development is not completely predictable given S.

The existence of a domain of dependence induces another definition concerning the boundaries of those regions as follows:

Definition 3.3.3 (Cauchy horizon). (Adapted from [72]) The future, past, or total Cauchy horizon of a closed set S of chronologically disconnected points is defined as (respectively):

- $H^{+}(S) = D^{+}(S) - I^{-}\left(D^{+}(S)\right)$.
- $H^{-}(S) = D^{-}(S) - I^{+}\left(D^{-}(S)\right)$.
- $H(S) = H^{+}(S) \cup H^{-}(S)$.

Given initial data on a set S of chronologically disconnected points on a hypersurface Σ, anything that happens in a region outside its Cauchy horizon is not completely predictable by using only that initial data. An illustration of both, domains of dependence and Cauchy horizons is presented in figure 3.8.

For geodesic motion in General Relativity, for each conserved quantity there is a vector field $\xi \in \Xi(\mathcal{M})$ such that

$$\mathcal{L}_{\xi}\mathbf{g} = 0, \tag{3.22}$$

called Killing vector field. Each spacetime can have different symmetries and conserved quantities. Given a Killing vector field ξ and a vector field U that is tangent to a geodesic affinely parametrized by λ, one can show that [66]

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}\mathbf{g}(U,\xi) = 0. \tag{3.23}$$

Therefore, $\mathbf{g}(U,\xi)$ is a conserved quantity along the geodesic to which U is tangent.

Interpreting conserved quantities in order to identify them with constants of motion from flat spacetime physics like energy and momentum is not trivial, and each case should be analyzed separately. When working with an asymptotically flat spacetime,i.e., a spacetime that becomes Minkowski for an observer situated far away from the gravitational source, one can evaluate the conserved quantity at infinity(far away from the gravitational source) and try to relate it with some known conserved quantity in flat spacetime.

Generally, a timelike Killing vector field results in a conserved quantity related to energy, when the considered geodesic is timelike. Because of that, a translation on the direction of a timelike Killing vector field preserves energy, like time evolution in flat spacetime. Then, when available, one defines a natural time direction as being parallel to that of the timelike Killing vector field. When there is no timelike Killing vector field, it is not possible to set up a natural definition of time. Therefore, timelike Killing vector fields are crucial to canonically quantize fields on curved spacetimes.

In this text we will be working exclusively with asymptotically flat spacetimes, hence, at least in asymptotic regions, there will be a timelike Killing vector field, allowing one to quantize a field without arbitrariness on the choice of a notion of time.

3.5 Black hole thermodynamics

Black holes obey a set of specific laws relating their properties like mass and area. Those laws are known as the laws of Black hole mechanics and sometimes as the laws of Black hole thermodynamics due to the resemblance that they have with the laws of thermodynamics presented before in 2.1. Up to this day, the equivalence between the laws of Black hole mechanics and thermodynamics is not generically proven and still a conjecture. In this section, these laws will be presented without their derivations. For the derivations, the reader is referred to [75] and [66]. For a textbook on Black hole thermodynamics, the reader is referred to [76] and for reviews on that topic, the reader is referred to [77, 78, 79].

All four laws of Black hole thermodynamics are valid for an asymptotically flat stationary Black hole. The uniqueness theorems for Black holes [78] guarantees that the most general stationary and asymptotically flat Black hole solution for Einstein-Maxwell equation is the Kerr-Newman Black hole[80, 81]. A Black hole of that kind with mass M, electric charge Q and angular momentum J is described, in Boyer-Lindquist coordinates, by the following metric

$$\begin{cases} \mathrm{d}s^2 = -\left(\frac{\mathrm{d}r^2}{\Delta} + \mathrm{d}\theta^2\right)\rho^2 + \left(c\,\mathrm{d}t - a\sin(\theta)^2\,\mathrm{d}\phi\right)^2\frac{\Delta}{\rho^2} - \left[\left(r^2+a^2\right)\mathrm{d}\phi - ac\,\mathrm{d}t\right]^2\frac{\sin(\theta)^2}{\rho^2} \\ t \in \mathbb{R},\ r \in \mathbb{R}_{>0},\ \theta \in [0,\pi]\ ,\phi \in [0,2\pi[\,, \end{cases} \tag{3.24}$$

where $a = \frac{J}{Mc}$, $\rho^2 = r^2 + a^2\cos(\theta)^2$, $\Delta = r^2 - r_s r + a^2 + {r_Q}^2$, ${r_Q}^2 = \frac{Q^2 G}{4\pi\epsilon_0 c^4}$, r_s is the Schwarzschild radius and ϵ_0 is the vacuum permittivity. We now proceed to the statement of the laws of Black hole mechanics.

Zeroth law of Black hole mechanics 3.5.1. (Adapted from [66, 77, 79]) The surface gravity of a Black hole is uniform over the entire event horizon if one or more of the following conditions are satisfied:

1. The Black hole is static.
2. The Black hole is stationary-asymmetric with a $t \to -t, \phi \to -\phi$ reflection symmetry of the time and angular coordinates.
3. The Black hole spacetime obeys Einstein's equation with the matter stress-energy tensor satisfying the dominant energy condition.

To properly understand the zeroth law of Black hole mechanics, it is necessary to introduce the new terms presented on its statements: static and stationary spacetimes, and the dominant energy condition. A spacetime $(\mathcal{M}, \mathbf{g})$ is said to be static if its metric $\mathbf{g}$ admits a timelike Killing vector field $K \in \Xi(\mathcal{M})$ which is orthogonal to a family of spacelike surfaces[4]. When a spacetime is static, one can always find a metric without dependence on the timelike coordinate.

An asymptotically flat spacetime $(\mathcal{M}, \mathbf{g})$ is said to be stationary if there exists, on $\mathcal{M}$, a complete Killing vector field which is timelike in the asymptotic region [82].

Finally, the matter stress-energy tensor $T_{\mu\nu}$ satisfies the dominant energy condition when for every timelike $W \in \Xi(\mathcal{M})$, $T^{\mu\nu}W_{\mu}W_{\nu} \geq 0$, and $T^{\mu\nu}W_{\mu}$ is a non-spacelike vector[4]. When the dominant energy condition is satisfied, the local energy density is non-negative for any observer and the local energy flow vector $T^{\mu\nu}W_{\mu}$ is non-spacelike, in other words, the energy flows at most at the speed of light.

According to [71], section 9.3, the solution for Einstein's equations outside a collapsed object are expected to settle down to one of the Kerr family solutions[4] if the collapsing object was uncharged or one of the Kerr-Newman family solutions if that was charged. Therefore, the final state of the classical evolution of a Black hole spacetime is at most a Kerr-Newman Black hole.

The next idea is that of surface gravity, which is usually represented by κ and is the force required of an observer at infinity to hold a particle(of unit mass) in place at the event horizon [66]. According to the zeroth law, κ is constant over H^+ for a stationary Black hole.

First law of Black hole mechanics 3.5.1. (Adapted from [83] and [77]) The changes in mass M, surface area A, angular momentum J and electric charge Q of a stationary Black hole due to any nonsingular, asymptotically flat perturbation are related by

$$\delta M = \frac{\kappa}{8\pi G}\delta A + \frac{1}{c^2}\Omega_H \delta J_H - \frac{1}{c^2}\Phi\delta Q, \tag{3.25}$$

[4] Kerr-Newman with zero electric charge.

where κ, Ω_H and Φ are, respectively, the surface gravity, angular velocity and electric potential of the Black hole.

In this law, the mass, angular momentum and electric charge should be understood as the ones computed on the Arnowitt-Deser-Misner (ADM) formalism [84], and the area of the Black hole is defined as the area of the event horizon.

The first law applies to a situation where an otherwise stationary Black hole of mass M, surface area A, angular momentum J and charge Q is perturbed by a small quantity of matter described by the stress-energy tensor $T_{\alpha\beta}$ in a quasi-static process such that at the end the Black hole is brought into another stationary state with parameters $M + \delta M$, $J + \delta J$, $A + \delta A$ and $Q + \delta Q$ after a non-stationary transient state.

Second law of Black hole mechanics 3.5.1. (As is in [66]) If the null energy condition is satisfied, then the surface area of a Black hole can never decrease: $\delta A \geq 0$.

The stress-energy tensor $T_{\mu\nu}$ is said to satisfy the null energy condition if for any future-directed null vector field $k \in \Xi(\mathcal{M})$

$$T_{\mu\nu} k^\mu k^\nu \geq 0. \tag{3.26}$$

Third law of Black hole mechanics 3.5.1. (As is in [66]) If the stress-energy tensor is bounded and satisfies the weak energy condition, then the surface gravity of a Black hole cannot be reduced to zero within a finite advanced time.

The stress-energy tensor $T_{\mu\nu}$ is said to satisfy the weak energy condition if for any time-like vector field $k \in \Xi(\mathcal{M})$

$$T_{\mu\nu} k^\mu k^\nu \geq 0. \tag{3.27}$$

Given the four laws of Black hole mechanics, one can make the analogy with the laws of thermodynamics by identifying the thermodynamic entropy S with the Black hole area A(up to a constant that corrects the dimension), as both are quantities that either reduce or stay constant in the case of isolated systems. With this identification, one can resort to the first law of Black hole mechanics and identify, up to a dimensional constant, κ as the thermodynamic temperature T which is the generalized force associated with a variation of the entropy. Hence, the laws of Black hole mechanics seem to be a statement of the laws of thermodynamics. Bekenstein initially proposed that analogy in [2], and the main counter-argument was that if a Black hole has a thermodynamic temperature, it then must emit thermal radiation, and Black holes were thought to be objects such that anything that enters it cannot escape and because of that, it is not expected to be a radiation emitter. Hawking concluded [3] that Black holes could indeed radiate if one considers a quantum field in a classical spacetime background. Hence, at least theoretically, Black holes do radiate. The radiation from stationary Black holes is called Hawking radiation, and the Black hole entropy is called Bekenstein-Hawking entropy.

The entropy and temperature of a Schwarzschild Black hole are given by, respectively,

$$S_{BH} = k_B \frac{c^3}{4\hbar G} A, \tag{3.28}$$

$$T_H = \frac{\hbar c^3}{8\pi G k_B} \frac{1}{M}. \tag{3.29}$$

These quantities result from a semi-classical treatment of a quantum field on a curved spacetime background. The presence of the Planck constant is taken as evidence of the underlying quantum phenomena producing such entropy and temperature.

There is still no proof that the entropy and temperature of a Black hole are of a thermodynamic nature. Up to now there is no satisfactory statistical mechanical treatment for Black holes that results in the laws of Black hole mechanics. This criterion arises from the fact that thermodynamics arises from a statistical treatment of either classical or quantum mechanics. For a review on the approaches to do statistical mechanics of Black holes, the reader is referred to [85].

3.6 The semi-classical framework for Gravity

To this time there is no complete quantum theory for gravity. As an approximation, one resorts to the semi-classical framework where the gravitational field is treated classically, and matter fields are treated as quantum fields. While the idea is simple, the actual calculations are involved.

The starting point is Einstein's equation

$$G_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu}\,, \tag{3.30}$$

where the stress-energy tensor is calculated using Hilbert's definition,

$$T_{\mu\nu} = -\frac{2}{\sqrt{-g}} \frac{\delta S}{\delta g^{\mu\nu}}, \tag{3.31}$$

and S is the action of the underlying theory. One then solves Einstein's equations to obtain the resulting spacetime metric $\mathbf{g}$.

In the semi-classical framework, one adds the renormalized stress-energy tensor $\langle\psi|\,\hat{T}_{\mu\nu}\,|\psi\rangle_{ren}$ to Einsetein's equation, where $|\psi\rangle$ is the quantum state of a matter field. The resulting semi-classical equation is:

$$G_{\mu\nu} = \frac{8\pi G}{c^4}\left(T_{\mu\nu} + \langle\psi|\,\hat{T}_{\mu\nu}\,|\psi\rangle_{ren}\right). \tag{3.32}$$

Then, by solving for the metric $\mathbf{g}$, one gets the self-consistent solution of the semi-classical equation. Unfortunately, exactly solving that equation is very difficult and one appeals to a perturbative solution. One begin the pertubative approach by considering only the classical stress-energy tensor and solving for the metric $\mathbf{g}^{(0)}$, which is interpreted as the zeroth order approximation to the self-consistent solution. For the next step, one evaluate $\langle\psi|\,\hat{T}_{\mu\nu}\left(\mathbf{g}^{(0)}\right)|\psi\rangle_{ren}$ and plug it into the semi-classical equation:

$$G_{\mu\nu}\left(\mathbf{g}^{(1)}\right) = \frac{8\pi G}{c^4}\left(T_{\mu\nu}\left(\mathbf{g}^{(0)}\right) + \langle\psi|\,\hat{T}_{\mu\nu}\left(\mathbf{g}^{(0)}\right)|\psi\rangle_{ren}\right). \tag{3.33}$$

By solving for $\mathbf{g}^{(1)}$, one gets the first order approximation to the self-consistent solution. By repeating this process one gets higher order approximations of the self-consistent solution.

After the semi-classical equations are solved, one gets a backreacted spacetime $(\mathcal{M}', \mathbf{g}')$ that is different from the classical spacetime $(\mathcal{M}, \mathbf{g})$. An speculative example of a backreacted spacetime resulting from semi-classical equations is presented in figure 3.9.

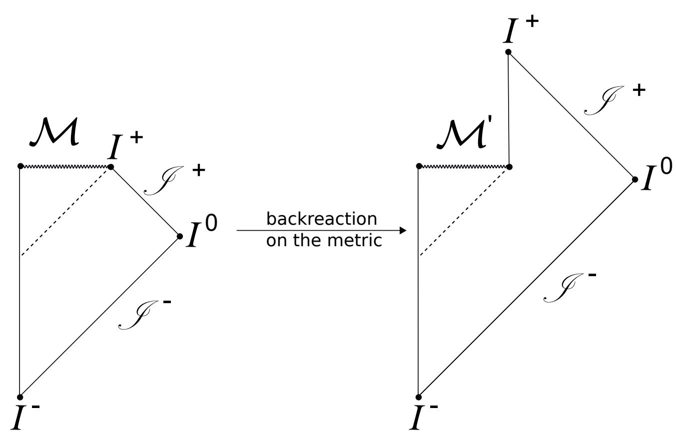

Figure 3.9: The classical spacetime $(\mathcal{M}, \mathbf{g})$ resulting from matter collapse and its backreacted version $(\mathcal{M}', \mathbf{g}')$ on a complete evaporation scenario. We remark that this is an speculative scenario that is widely used for constructing arguments on the information loss puzzle.

4

The information loss paradox

Why is information loss paradoxical in the first place? Why can't we simply let information be lost and assume everything is alright? Where is the contradiction that supports the paradox? Why can't information be held in a remnant, i.e., in an object remaining from Black hole evaporation? This chapter is dedicated to review these questions based on the available literature.

4.1 The paradox

To answer the first question posed in the introduction of this chapter, two paradoxes will be presented in separate subsections[1], each one concerning a different contradiction.

As a common starting point, let $\mathcal{M}$ be the spacetime representing the collapse of some classical matter to form a Black hole. At early times we consider that matter to be so diffuse that the spacetime is nearly flat. We then take $\mathcal{M}$ to be the classical background spacetime on which we consider an arbitrary quantum field[2] ϕ. We assume that at early times $\mathscr{I}^-$, the field ϕ is in the Minkowski vacuum state, which is a pure quantum state by definition.

4.1.1 Unitarity violation

Being the simplest statement of the information loss paradox, it was noted by Hawking as follows [4] : Considering the semi-classical framework, the state of ϕ on $\mathscr{I}^+$ will be thermal with emission spectrum equivalent to that of a gray body[3] at temperature $\dfrac{\kappa}{2\pi}$. Then, a pure state has evolved into a mixed state. Hence, the evolution is non-unitary. An illustration of this statement for collapsing null matter, i.e., radiation, can be seen in figure 4.1. We enforce that the paradox would still hold had we considered massive matter, as in the case of figure 4.3.

To conclude that the emitted radiation is exactly thermal an approximation is used [86]. More precisely, the dominant contributions to the state of the field on $\mathscr{I}^+$ come from

[1] Despite being different paradoxes, both are called information loss paradox.
[2] For simplicity, all indexes, i.e., vector, spinor, etc, were omitted.
[3] A gray body spectrum is proportional to a black body spectrum.

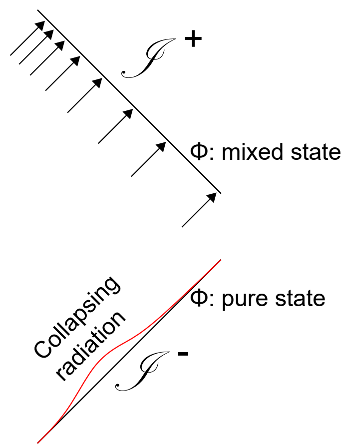

Figure 4.1: On $\mathscr{I}^-$ there is collapsing radiation and a quantum field in Minkowski vacuum state. On $\mathscr{I}^+$ there is a quantum field in a thermal state. The arrows represent radiation going to $\mathscr{I}^+$.

late retarded times and contributions from early retarded times are negligible compared to them. Physically, the neglected contributions are equivalent to a finite amount of radiation emitted by the collapsing matter prior to the formation of the Black hole. The radiation emitted during that intermediary phase does depend on the details of the collapse. But the particles emitted during that state are so disperse that their contribution to the emission spectrum at $\mathscr{I}^+$ is negligible. Therefore, there is no reason to think that unitarity would be restored by considering that intermediary emission.

4.1.2 The Bekenstein bound

The next statement of the information loss puzzle is based on Page's paper [87]. From now on we will make use of horizon intersecting hypersurfaces whose definition is introduced below:

Definition 4.1.1 (Horizon intersecting hypersurface). A hypersurface Σ is a horizon intersecting hypersurface if $\Sigma \cap \partial\mathcal{B} \neq \varnothing$, where $\mathcal{B}$ is the Black hole region of a spacetime.

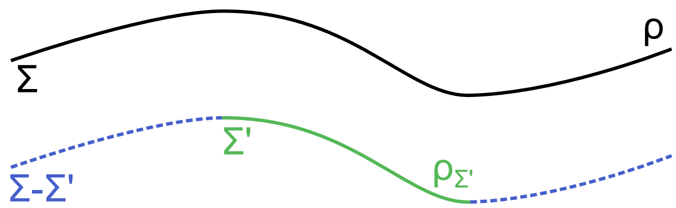

Figure 4.2: In black is the spacelike hypersurface Σ where the state ρ of the quantum field is defined. In green is another spacelike hypersurface $\Sigma' \subset \Sigma$. By tracing out quantum states defined on $\Sigma - \Sigma'$, represented by the blue dashed line, one gets the quantum state $\rho_{\Sigma'}$, which describes measurements done by someone with access only to data on Σ'.

Before proceeding to the statement, we make a brief pause to present a notation we will be using throughout the text. Let a quantum field ϕ be in state ρ on a spacelike hypersurface Σ. Now consider another spacelike hypersurface[4] $\Sigma' \subset \Sigma$. Any measurement that is done on ϕ by an observer with access only to information on Σ' is described by the following density operator:

$$\rho_{\Sigma'} = \mathrm{Tr}_{\Sigma - \Sigma'}\, \rho. \tag{4.1}$$

That notation means that one takes the partial trace over states that are defined on $\Sigma - \Sigma'$. An illustration is presented in figure 4.2. Now, we proceed to the statement of the paradox:

1. The Black hole region is $\mathcal{B} = \mathcal{M}' - I^{-}\left(\mathscr{I}^{+}\right)$ and the rest $\mathcal{U}$ of the universe is $\mathcal{U} = \mathcal{M}' - \mathcal{B}$, i.e., everything outside the Black hole.

2. Consider a complete spacelike horizon intersecting hypersurface Σ, corresponding to an instant of time after the Black hole has formed. The Bekenstein-Hawking entropy[5] of the Black hole is $S_B(\Sigma) > 0$. Assuming that the state ρ of ϕ on Σ is pure, one can take the partial trace of ρ with respect to $\Gamma = \Sigma \cap \mathcal{U}$ to get the density op-

[4] On this text, $\subset$ is used as the symbol for proper subset.

[5] The reader should understand it as the Bekenstein-Hawking entropy of a Black hole with a surface area equal to the area of $\Sigma \cap \partial\mathcal{B}$.

erator $\rho_\Gamma = \mathrm{Tr}_{\Sigma-\Gamma}\,\rho$ that describes measurements done on ϕ by an observer outside the Black hole. The Von Neumann entropy of ρ_Γ is $S\left(\Gamma\right)$[6]. Furthermore, the Von Neumann entropy of the inaccessible portion of ϕ[7] is equal to that of the accessible portion, i.e., $S\left(\Gamma\right) = S\left(\Sigma - \Gamma\right)$ as they compose a bi-partition of the pure state ρ of ϕ, in other words, $S\left(\Gamma\right)$ and $S\left(\Sigma - \Gamma\right)$ are entanglement entropies.

3. Let $\Sigma' \subset I^+\left(\Sigma\right)$ be a complete spacelike horizon intersecting hypersurface contained in the chronological future of Σ. At the instant of time corresponding to Σ', $S_B\left(\Sigma'\right) < S_B\left(\Sigma\right)$ as there is a negative energy flux in across the event horizon due to the emission of Hawking-Radiation. The portion of Σ' situated outside the Black hole is $\Gamma' = \Sigma' \cap \mathcal{U}$. The state of ϕ outside the event horizon is $\rho_{\Gamma'}$, with Von Neumann entropy $S\left(\Gamma'\right) > S\left(\Gamma\right)$ as the state of ϕ outside the Black hole is expected to approach a thermal state as Γ' approaches $\mathscr{I}^+$. Moreover, the Von Neumann entropy of the other part of ϕ is $S\left(\Sigma' - \Gamma'\right) = S\left(\Gamma'\right)$.

4. Then, the Von Neumann entropy of the portions of ϕ both inside and outside the Black hole are increasing during the evaporation process due to the emission of Hawking radiation by the Black hole. On the other hand, the Bekenstein-Hawking entropy of the Black hole is decreasing due to that emission. At some instant of time, the Bekenstein-Hawking entropy of the Black hole will be smaller than the entanglement entropy between the inside and outside parts of ϕ. That instant of time is called the Page time. By assuming that the Bekenstein-Hawking entropy is the upper bound for the Von Neumann entropy of a region of spacetime, a paradox arises as the entan-

[6]This is to be understood as the Von Neumann entropy of the state ρ_Γ of the quantum field ϕ over Γ.
[7]According to an observer outside the Black hole.

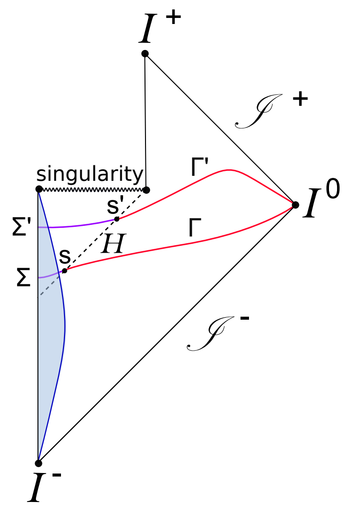

Figure 4.3: In the drawing the hypersurfaces Σ and Σ' are shown in purple and red. In red, $\Gamma = \Sigma - \mathcal{B}$ and $\Gamma' = \Sigma' - \mathcal{B}$. In purple, $\Sigma - \Gamma$ and $\Sigma' - \Gamma'$. The horizon surfaces S and S' at the respective instants of time Σ and Σ' are marked by dots. In blue is the collapsing matter. We warn the reader that, while appearing several times in the literature, this conformal diagram should not be taken as a the literal backreacted spacetime but only as a tool to reason about Black hole evaporation. This warning is valid whenever this diagram is used in this text.

glement entropy will be bigger than the Bekenstein-Hawking entropy, violating the upper bound of Von Neumann entropy.

The situation is illustrated in figure 4.3.

The assumption of an upper bound for the entropy, used in point 4 above, is backed by Bekenstein's entropy bound [88], which was generalized by Bousso as the covariant entropy conjecture [89]. For reviews on those entropy bounds, the reader is referred to [90] and [91]. Furthermore, by assuming that the Bekenstein-Hawking entropy is the thermodynamic entropy of the Black hole, one is implicitly assuming that it is also an upper bound for its Von Neumann entropy. That happens because the thermodynamic entropy is defined as the maximum Von Neumann entropy, up to Boltzmann's constant, for a given macrostate of a physical system.

The importance of this statement is that the paradox arises at scales where the semi-classical framework is expected to hold. It is important to remark that this statement of the paradox relies on the assumption that ϕ evolves unitarily. By dropping this assumption, the notion of entanglement entropy is lost and the statement does not hold.

4.2 The BSP argument

In principle, one could argue that there is no physical problem with non-unitary evolution. Following this point of view, Hawking proposed the superscattering [4] approach, which consisted of a generalization of the scattering operator(S-matrix) that allows pure states to evolve into mixed states. However, a strong counterargument was given by Banks, Susskind & Peskin[5]. They showed that such approach would violate either locality[8] (by creating observable correlations between widely separated points on the same spacelike hypersurface) or energy-momentum conservation. None of these violations is desirable due to the catastrophic physical consequences like faster-than-light communication or unbounded violation of energy conservation.

According to [93], if the dynamics which gives rise to the superscaterring operator is local in time, it can be represented as the integral of the following differential equation:

$$\dot{\rho} \equiv \frac{\mathrm{d}\rho}{\mathrm{d}t} = \not{H}\rho, \tag{4.2}$$

where t is the time and $\not{H}$ is an arbitrary linear operator constrained to preserve hermiticity, positivity and normalization of the density operator ρ. This equation can be cast into the

[8]The notion of locality the authors were referring to is the one introduced by the cluster decomposition principle [92]

following form[5]:

$$\dot{\rho} = -i\left[H_0, \rho\right] - \frac{1}{2}\sum_{\alpha,\beta\neq 0}\left(Q^\beta Q^\alpha \rho + \rho Q^\beta Q^\alpha - 2Q^\alpha \rho Q^\beta\right), \tag{4.3}$$

where H_0 is an hermitian operator and $\{Q^\alpha\}_{\alpha=0}^N$ is a complete set of hermitian matrices that decomposes $\not{H}$, such that $Q^0 = \mathbf{1}$ and

$$\not{H}\rho = -\sum_{\alpha,\beta} h_{\alpha\beta} Q^\alpha \rho Q^\beta, \tag{4.4}$$

where $h_{\alpha\beta}$ are the expansion coefficients. Then,according to [5], under ordinary Quantum Mechanics, a system evolving under the following Hamiltonian

$$H\left(t\right) = H_0 + \sum_\alpha j_\alpha\left(t\right) Q^\alpha, \tag{4.5}$$

where j_α are c-number sources randomly varying in time according to gaussian statistics with covariance

$$\left\langle j_\alpha\left(t\right) j_\beta\left(t'\right)\right\rangle = h_{\alpha\beta}\delta\left(t - t'\right), \tag{4.6}$$

will obey evolution equation (4.3). However, the non-trivial time dependence of j_α implies that energy can be added or removed from the system, resulting in a violation of energy conservation.

When considering Quantum Field Theory, the former Hamiltonian must be generalized to:

$$H = H_0 + \int \mathrm{d}^3x\, j_\alpha\left(t, \vec{x}\right) Q^\alpha\left(\vec{x}\right), \tag{4.7}$$

where t is the time. If the sources fluctuate randomly with the spatial position, they will break translational invariance, allowing momentum to be added or removed from the system. On the other hand, if the fluctuations of that source were translationally invariant, the sources would go through the same fluctuations at widely separated points on the same spacelike hypersurface. That would introduce correlations between observables at spacelike separated points, thus violating locality.

The BSP argument is a very strong argument against non-unitary evolution. While very strong, it still have caveats that are explored by Unruh & Wald, whose arguments will be presented in the next section.

4.3 The UW arguments

Unruh & Wald presented a review [94] on the information loss puzzle and among other topics, presented arguments in favor of information loss. According to them, the main arguments against information loss are based on the following statements:

1. Information loss implies violation of unitary.
2. Information loss implies failure of energy conservation.
3. Information loss violates the AdS/CFT correspondence.

The counterarguments presented by them will be covered in separate subsections.

4.3.1 Information loss implies violation of unitary

This point was stated in [95] and received a didactic refinement in the review [94]. When one thinks about unitarity in the context of the information loss puzzle, there are two prin-

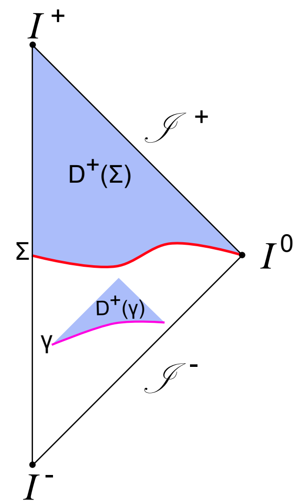

Figure 4.4: Σ is a Cauchy surface and its future Cauchy development covers the whole upper "half" of $\mathcal{M}$ and the past Cauchy development, if drawn, would cover the other "half". γ is a partial Cauchy surface and its future Cauchy development is much smaller than $\mathcal{M}$.

ciples that come into mind:

1. Conservation of probabilities.

2. Evolution from pure states into pure states.

According to the semi-classical picture, it is principle 2 that is being violated but not principle 1. But a violation of principle 2 is expected to happen whenever one chooses a surface of constant "time" that is not a Cauchy surface. To picture that argument, one can think of a Cauchy surface as a surface containing the initial data for an initial value problem in general relativity. Because of that, one expects the whole spacetime manifold $\mathcal{M}$ to be equal to the Cauchy development of a Cauchy surface. Examples of both Cauchy surfaces and partial Cauchy surfaces for the Minkowski spacetime are presented in figure 4.4.

The behavior of a matter field at any point $p \in \mathcal{M}$ that lies outside the Cauchy development of a partial Cauchy surface Γ is not completely predictable if that hypersurface is

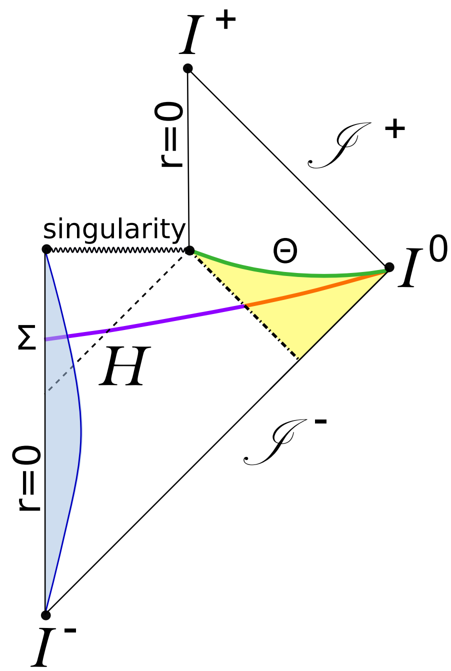

Figure 4.5: A spacetime resulting from the complete evaporation of a Black hole formed by the collapse of matter. In light blue is the matter that collapsed into the Black hole whose horizon H is the dashed line. The part of the past Cauchy horizon of $\mathscr{I}^+$ that has no overlapping with $\mathscr{I}^-$ is represented by the dot-dashed line. Σ is a spacelike hypersurface represented partly by a red line and partly by a purple line. Θ is another spacelike hypersurface represented by a green line. The past Cauchy development of Θ is the yellow shaded area.

given as initial data. What happens in the complete evaporation of a Black hole is that after evaporation there is no hypersurface such that its past Cauchy development covers the whole $\mathscr{I}^-$. The situation is depicted in figure 4.5. By looking at it one can see that given data on the hypersurface Σ it is possible to tell what will go on in Θ, as Θ is contained in the future Cauchy development of Σ. On the other hand, only part of Σ is contained in the past Cauchy development of Θ. Anything outside that region is not completely known by someone with access only to Θ. Hence, once the Black hole completely evaporates, part of the information about its formation and development is lost.

Therefore, an otherwise pure quantum state on Σ would be described as a mixed state on Θ, violating the second notion of unitarity. Hence, the information loss is tied to the absence of a Cauchy surface after complete evaporation and not to a fundamental violation of quantum theory itself. Because of that, there is no reason to modify the quantum theory in regimes away from Planck scale just to avoid such non-unitary evolution from happening. Then, according to Unruh & Wald, an otherwise pure quantum state on $\mathcal{I}^-$ ends up as a mixed state on $\mathscr{I}^+$

4.3.2 Information loss implies failure of energy conservation

The argument sustaining the association of the evolution of pure states into mixed states and failure of energy conservation is the famous BSP argument. To counter that argument, Unruh and Wald provided a class of models for evolution of pure states into mixed states [95] and constructed their superscattering operators, based on the generic model provided in [5]. With this they showed that the energy violation in that kind of model can be controlled, in the sense that such violations can be confined to Planck scale.

Moreover, Unruh and Wald argue that the class of models presented in the BSP argument are Markovian(local in time) and, therefore, the BSP argument has no effect on non-Markovian models. They provide [95] a non-Markovian toy model to illustrate their point. And also argue that Black holes should have a long-time-scale "memory" stored in its external gravitational field of the amount of energy that went into it and because of that one cannot expect an effective model of Black hole formation and evaporation to be Markovian in nature.

Furthermore, they argue that the widespread belief that quantum decoherence requires energy exchange is not true when the environment is taken to be a spin bath where excitation of the degrees of freedom of the environmental system does not require energy. Based on that idea Unruh provided a toy model of decoherence without dissipation in [96].

4.3.3 Information loss violates the AdS/CFT correspondence

In a few words, the AdS/CFT conjecture [97] says that a gravitational theory in an d-dimensional Anti-de Sitter (AdS) spacetime can be mapped to a Conformal Field Theory[9] (CFT) in $(d-1)$-dimensions defined on the boundary of that AdS space. Using that conjecture, one can argue that a theory of Quantum Gravity should be unitary as it is dual to a CFT which is known to be unitary, then, there must be a unitary description of Black hole evaporation, where there is no loss of information.

The critics Unruh&Wald make about the argument presented above, are that the AdS/CFT conjecture is still not developed to sufficient detail in order to furnish a precise and clear argument against information loss. They say that a more developed version of the conjecture should explain when and where is the semi-classical framework violated and how information is regained, rather than making the claim presented in the last paragraph.

4.4 Black Hole complementarity

In an attempt to reconcile Black hole evaporation with quantum mechanics, Susskind, Thorlacius & Uglum introduced [98] the Black hole complementarity hypothesis, sharing a viewpoint similar to the one presented by 't Hooft [99]. It consists of 4 postulates[10]:

[9] In other words, a Quantum Field Theory with Conformal Invariance

[10] As pointed in [9], postulate 4 is stated as a certainty in [98].

- Postulate 1: The process of formation and evaporation of a Black hole, as viewed by a distant observer, can be described entirely within the context of standard quantum theory. In particular, there exists a unitary S-matrix which describes the evolution from infalling matter to outgoing Hawking-like radiation[11].

- Postulate 2: Outside the stretched horizon(defined below) of a massive Black hole, physics can be described to a good approximation by a set of semiclassical field equations.

- Postulate 3: To a distant observer, a Black hole appears to be a quantum system with discrete energy levels. The dimension of the subspace of states describing a Black hole of mass M is the exponential of the Bekenstein entropy $S_B(M)$ [100].

- Postulate 4: A freely falling observer experiences nothing out of the ordinary when crossing the horizon.

To understand the consequences of those postulates, consider a spacetime $\mathcal{M}$ representing the collapse of some quantum matter described by a quantum field ϕ in a pure state $\left|\phi\left(\mathscr{I}^-\right)\right\rangle$ at $\mathscr{I}^-$ to form a Black hole. Assume that the Black hole evaporates completely to avoid a remnant scenario. On $\mathcal{M}$, consider two observers, Enzo and Valentina, who follow the word lines γ_E and γ_V respectively. An illustration of that setup is presented in figure 4.6.

According to postulate 1, any observer far from the Black hole will describe the whole situation within the context of standard quantum theory. We can then suppose that the state of ϕ, which describes measurements done by those far observers, is such that for two

[11]Hawking-like because that radiation does carry information.

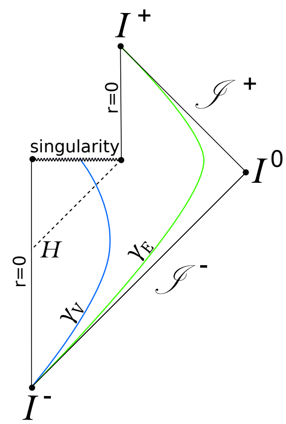

Figure 4.6: The dashed line represents the event horizon H of the Black hole. In blue is Valentina's worldline γ_V and in green in Enzon's worldline γ_E.

spacelike hypersurfaces Σ_1 and $\Sigma_2 \subset J^+(\Sigma_1)$, corresponding to two instants of time, there exists a unitary operator $U(\Sigma_1, \Sigma_2)$ such that

$$|\phi(\Sigma_2)\rangle = U(\Sigma_1, \Sigma_2)\, |\phi(\Sigma_1)\rangle\,. \tag{4.8}$$

That is illustrated in figure 4.7. The, the state of ϕ,on which Enzo makes his measurements, at $\mathscr{I}^+$ is

$$\left|\phi\left(\mathscr{I}^+\right)\right\rangle = U\left(\mathscr{I}^-, \mathscr{I}^+\right)\left|\phi\left(\mathscr{I}^-\right)\right\rangle, \tag{4.9}$$

which is a pure state. The existence of such time evolution would solve the information loss puzzle for Enzo or any other far observer.

Now let a spacelike hypersurface Σ represent a region of spacetime corresponding to an instant of time after Black hole evaporation, i.e., $\left(\partial D^-(\Sigma) \cap \mathscr{I}^-\right) \neq \varnothing$ and $\left(\partial D^-(\Sigma) \cap \mathscr{I}^-\right) \subset \mathscr{I}^-$. An example is given in figure 4.8. The state of ϕ which describes measurements done

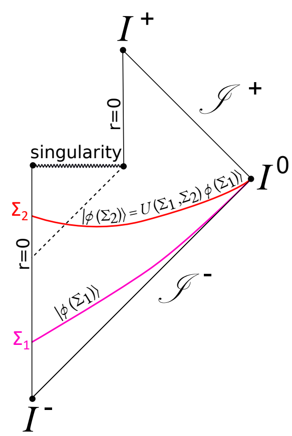

Figure 4.7: The dashed line represents the event horizon H of the Black hole. In pink is the spacelike hypersurface Σ_1 where the state of the quantum field is $|\phi(\Sigma_1)\rangle\rangle$. In red is the spacelike hypersurface Σ_2. The state of the quantum field on Σ_2 is related to the state of that field on Σ_1 by equation (4.8).

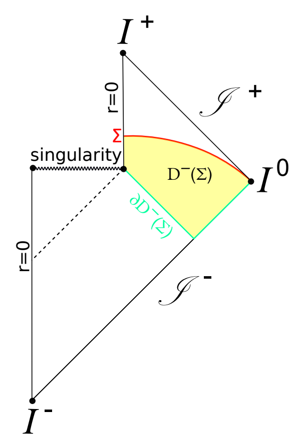

Figure 4.8: In red is a spacelike hypersurface Σ representing a region of spacetime corresponding to an instant of time after Black hole evaporation. In yellow is its past Cauchy development $D^-(\Sigma)$, and in light blue is the boundary of its past Cauchy development $\partial D^-(\Sigma)$. In this picture, it is easy to see that $\partial D^-(\Sigma)$ has a non-empty intersection with $\mathscr{I}^-$.

by distant observers at the instant of time corresponding to Σ is pure and given by

$$|\phi(\Sigma)\rangle = U\left(\mathscr{I}^-, \Sigma\right) \left|\phi\left(\mathscr{I}^-\right)\right\rangle . \tag{4.10}$$

As the time evolution operator is unitary, the initial state can be written as

$$\left|\phi\left(\mathscr{I}^-\right)\right\rangle = U\left(\mathscr{I}^-, \Sigma\right)^\dagger |\phi(\Sigma)\rangle . \tag{4.11}$$

Then, all quantum information that was encoded in $\left|\phi\left(\mathscr{I}^-\right)\right\rangle$ must also be encoded in $|\phi(\Sigma)\rangle$. This is the first consequence of Black hole complementarity: Some mechanism transfers all the quantum information that entered in the Black hole to the region $D^-(\Sigma)$, otherwise it would not be possible to write $\left|\phi\left(\mathscr{I}^-\right)\right\rangle$ as an invertible function of $|\phi(\Sigma)\rangle$.

The mechanism responsible for storing and transferring information about what fell into the Black hole is the so-called stretched horizon $\mathcal{H}_s$, a membrane situated a Planck length outside the event horizon $\mathcal{H}$ that describes a quasi-stationary Black hole. The set $\mathcal{H}_s$ is defined [98] by shifting every point $y \in \mathcal{H}$ along a past directed null curve by an arbitrarily small constant amount δ in such a way that

$$A(x) = A(y) + \delta, \tag{4.12}$$

where $x \in \mathcal{H}_s$ is a 2-dimensional surface and $A(p)$ is its area. The stretched horizon is a physical object, i.e., it has physical properties like electrical resistivity and viscosity, and plays the role of the quantum system from postulate 3. It stores all the information about matter crossing the event horizon and emits it back in the form of Hawking-like radiation.

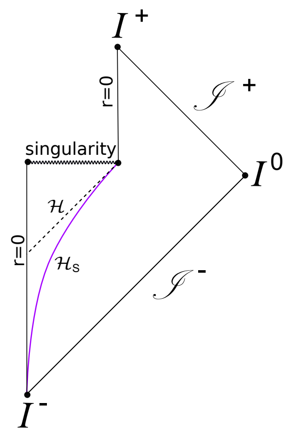

Figure 4.9: The dashed line represents the event horizon $\mathcal{H}$. In purple is the stretched horizon $\mathcal{H}_s$. Note that in a realistic representation, the stretched horizon would almost overlap the event horizon, in this illustration we exaggerated the distance between the horizons for a pedagogical reason.

An example of a stretched horizon is depicted in figure 4.9.

That mechanism has an interesting consequence. To understand it, we will analyze the perspective of Valentina, the infalling observer. Her measurements will agree with Enzo's measurements until she crosses the stretched horizon. According to postulate 2, after crossing that horizon, semiclassical equations are not guaranteed to be a good approximation of the situation. Moreover, postulates 1 and 3 do not cover the region past the event horizon. In a more precise statement, postulate 2 do not cover any instant of time corresponding to a spacelike hypersurface Λ_1 such that $\Lambda_1 \cap \gamma_V \subset I^+(\mathcal{H}_s \cap \gamma_V)$. Furthermore, postulates 1 and 3 fail at any instant of time corresponding to a spacelike hypersurface Λ_2 such that $\Lambda_2 \cap \gamma_V \subset I^+(\mathcal{H} \cap \gamma_V)$. Then, Enzo and Valentina agree about their measurements on ϕ at any instant of time corresponding to a spacelike hypersurface Λ such that $\Lambda \cap I^+(\mathcal{H}_s \cap \gamma_V) = \varnothing$. An illustration of those points is presented in figure 4.10

Given that there exists a region of disagreement between measurements done on ϕ by Enzo and Valentina, we will introduce a subscript $E(V)$ to indicate that a quantity de-

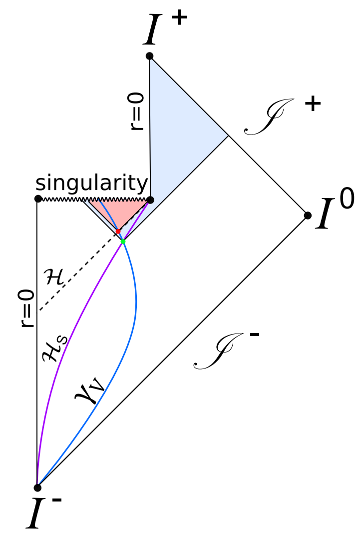

Figure 4.10: The dashed line represents the event horizon $\mathcal{H}$. In purple is the stretched horizon $\mathcal{H}_s$. The light blue line is Valentina's worldline γ_V. The green dot is $\mathcal{H}_s \cap \gamma_V$, where Valentina crosses the stretched horizion. The light blue region is $I^+\left(\mathcal{H}_s \cap \gamma_V\right)$. The red dot is $\mathcal{H} \cap \gamma_V$, where Valentina crosses the event horizon. The light red region is $I^+\left(\mathcal{H} \cap \gamma_V\right)$. The white region of the conformal diagram contains the spacelike hypersurfaces representing the instants of time when Valentina and Enzo agree about the state of ϕ.

scribes measurements done by Enzo(Valentina). Now consider a spacelike hypersurface Π such that $\Pi \not\subset J^-\left(\Lambda_1\right)$. The state of ϕ which contains information about measurements done by Enzo or any other far observers is pure and given by

$$\left|\phi\left(\Pi\right)\right\rangle_E = U\left(\mathscr{I}^-, \Pi\right)_E \left|\phi\left(\mathscr{I}^-\right)\right\rangle . \tag{4.13}$$

But if this is the case, no information about the quantum field ϕ crosses the horizon. The state of ϕ describing measurements done by far observers is uncorrelated to the state of ϕ inside the Black hole:

$$\left|\phi\left(\Pi\right)\right\rangle_E = \left|\phi\left(\Pi - \mathcal{B}\right)\right\rangle_E \otimes \left|\phi\left(\Pi \cap \mathcal{B}\right)\right\rangle_E , \tag{4.14}$$

where

$$|\phi\left(\Pi-\mathcal{B}\right)\rangle_E=\langle\phi\left(\Pi\right)|\,\mathrm{Tr}_{\mathcal{B}}\left(|\phi\left(\Pi\right)\rangle\langle\phi\left(\Pi\right)|_E\right)\tag{4.15}$$

$$|\phi\left(\Pi\cap\mathcal{B}\right)\rangle_E=\langle\phi\left(\Pi\right)|\,\mathrm{Tr}_{\Pi-\mathcal{B}}\left(|\phi\left(\Pi\right)\rangle\langle\phi\left(\Pi\right)|_E\right).\tag{4.16}$$

On the other hand, according to postulate 4, infalling observers should experience nothing out of the ordinary when crossing the event horizon, therefore, the state of ϕ that describes measurements done by Valentina or any other infalling observer is also pure and given by

$$|\phi\left(\Pi\right)\rangle_V=U\left(\mathscr{I}^-,\Pi\right)_V\left|\phi\left(\mathscr{I}^-\right)\right\rangle.\tag{4.17}$$

The infalling observers should freely pass through the stretched horizon, concluding that it does not exist, due to postulate 4. Therefore, the state of ϕ that describes measurements done by infalling observers presents correlations between the inside and outside portions of the field ϕ, implying that it cannot be decomposed as the product of an inside part and an outside part without losing some information. This conclusion rules out the possibility of defining a global quantum state for ϕ upon which any observer always agrees. The disagreement between measurements done by Enzo and Valentina on ϕ is not a problem because after crossing the event horizon, she has no way of sending Enzo any information about her observations, i.e., they cannot meet and compare notes after Valentina crosses the event horizon.

On top of that, there is another problem. Valentina brings information from outside with her, but Enzo may also get that information by making measurements on the radiation that originated on the stretched horizon. Hence, the membrane is cloning quantum information, which would violate the no-cloning theorem 4.4.1 of quantum mechanics, stated after this paragraph. It is argued that such a violation can never be witnessed because, as said before, they can not meet and compare notes after Valentina crosses the event horizon.

Theorem 4.4.1 (No-cloning theorem). (Adapted from [10, 101]) Given an arbitrary unknown quantum state $|\psi\rangle$, and two auxiliary known quantum states $|\phi\rangle$ and $|M\rangle$, define

$$|\Psi\rangle = |\psi\rangle \otimes |\phi\rangle \otimes |M\rangle \, . \tag{4.18}$$

Then, there is no unitary operator U such that

$$U\,|\Psi\rangle = |\psi\rangle \otimes |\psi\rangle \otimes |M\,(\Psi)\rangle \, . \tag{4.19}$$

If a quantum description of gravity imposes the existence of a global quantum state for any instant of time, the complementarity hypothesis is immediately falsified as noted in [98].

The term complementarity comes from the fact that any observer outside the stretched horizon can conduct experiments on its surface, probing properties such as electrical resistivity and viscosity. On the other hand, an observer that falls past the stretched horizon will note that it does not exist, it disappears as soon as the observer passes through it. However, there is no way of reporting the non-existence of the membrane to an observer outside the Black hole. Then, it is said that the observations made by infalling observers who cross the horizon and by distant observers are complementary. The choice falling or not inside the Black hole and its influence on the observed reality can be understood as an analogy to the influence of measuring or not on which slit a photon had passed on a double-slit experiment.

The arguments against the Black hole complementarity are

- Nice slice argument: Presented in [102], it consists in the construction of a family of Cauchy surfaces that foliate the geometry in such a way that the surfaces avoid regions of strong spacetime curvature and yet cut through the infalling matter and the outgoing Hawking radiation. Furthermore, those surfaces are required to be everywhere smooth, with small extrinsic curvature compared to any microscopic scale. Under such a setup it is argued that no significant deviation from the usual local quantum field theories is expected. Hence, The degrees of freedom outside the Black hole are expected to hold information about the degrees of freedom inside the Black hole as expected from local quantum field theories. In [102], a remark is presented where the authors say that the nice slice argument fails when one considers string theory.

- Black hole complementarity is unnecessary: The S-matrix ansatz[12] is complete by itself, that is, such assumption implies that strong gravitational interactions should take place just outside the horizon (in order to let information escape the black hole encoded in Hawking-like radiation) and those interactions result in a backreaction that is strong enough to keep infalling particles from crossing the event horizon. Hence, black hole complementarity is not neccessary. This argument was presented in [103]. Note that the S-matrix ansatz has an important additional requirement: It

[12] The assumption that the formation and evaporation of a Black hole can be described in terms of an S-matrix, which was introduced by [99].

requires a renormalization scheme for the stress-energy tensor that does not assumes regularity at the horizon. One should take that into account when thinking about this argument.

- Event horizons are not ordinary regions: Originally presented in [104], it is argued that according to string theory on the Euclidean version of the Schwarzschild Black hole, there is a zero mode at the horizon. Then, the Black hole event horizon is no ordinary region due to the presence of that zero-mode. Hence, according to this argument, the postulate 4 of the Black hole complementarity hypothesis is violated.

- Complementarity is not enough: Presented in [105], it is argued that near the Black hole horizon the spacetime is approximately Rindler. It is known that there is entanglement between both Rindler wedges and in this case, one of those wedges lies inside the event horizon. An infalling observer can collect information about the near-horizon Hawking radiation modes. By doing that he might be able to witness a violation of the monogamy of entanglement due the near-horizon region being maximally entangled with early radiation and the Black hole interior. Hence, complementarity is not enough.

4.5 Black hole remnants

If the Black holes do not evaporate completely, one can assume that the missing information is enclosed in a region of spacetime, therefore solving both information puzzles. The Black hole remnant is proposed to be that region of spacetime containing information about the state of matter that felt into the Black hole. That is the most conservative proposal as it does not demand a departure from locality nor unitarity, its only assumption is that some new phenomenon halts Hawking Evaporation at Planck scale, where the usual theories are not supposed to be valid.

There are several different proposals that we classify later under the remnant type. As a common setup, we consider a spacetime $\mathcal{M}$ representing the collapse of some quantum matter described by a quantum field ϕ in a pure state $\left|\phi\left(\mathscr{I}^-\right)\right\rangle$ at $\mathscr{I}^-$ to form a Black hole of mass M. Considering the semi-classical framework, at late stages of evaporation, i.e., late retarded times, the state of the field ϕ will be approximately thermal, with a Bekenstein-Hawking entropy $S_B\left(\phi\right) \ = \ 4\pi K_B\left(\frac{M}{M_{pl}}\right)^2$. For the remnant scenario to emerge, we consider that when the Black hole mass reaches Planck scale, some phenomenon halts the evaporation process, producing a remnant.

To understand the reasoning concerning remnants, consider a spacelike hypersurface Σ, such that $\Sigma \cap \mathcal{B} \neq \varnothing$, corresponding to an instant of time after Black hole formation. The state of ϕ at that instant of time is expected to be pure and given by

$$\left|\phi\left(\Sigma\right)\right\rangle = U\left(\mathscr{I}^-, \Sigma\right)\left|\phi\left(\mathscr{I}^-\right)\right\rangle, \tag{4.20}$$

where U is some unitary time evolution operator. An observer outside the Black hole does not have access to information about the state of the field ϕ inside the event horizon. Then, only part of the complete quantum state of ϕ is necessary to describe measurements done by observers outside the Black hole. That part is given by

$$\rho\left(\phi, \Sigma - \Sigma \cap \mathcal{B}\right) = \mathrm{Tr}_{\Sigma\cap\mathcal{B}}\left[\left|\phi\left(\Sigma\right)\right\rangle\!\left\langle\phi\left(\Sigma\right)\right|\right], \tag{4.21}$$

which, according to the semi-classical framework, approaches a thermal state as Σ approaches $\mathscr{I}^+$. Then, measurements done by observers at $\mathscr{I}^+$ are described by a thermal state. For that reason, the Von Neumann entropy of the quantum state of ϕ that describes measurements done by observers at $\mathscr{I}^+$ is taken as equal to the Bekenstein-Hawking entropy of the Black hole,

$$S\left(\phi, \Sigma - \Sigma \cap \mathcal{B}\right) = 4\pi K_B \left(\frac{M}{M_{pl}}\right)^2 . \tag{4.22}$$

And is also equal to the Von Neumann entropy of ϕ on $\Sigma \cap \mathcal{B}$

$$S\left(\phi, \Sigma - \Sigma \cap \mathcal{B}\right) = S\left(\phi, \Sigma \cap \mathcal{B}\right) = 4\pi K_B \left(\frac{M}{M_{pl}}\right)^2 , \tag{4.23}$$

because the pure state of ϕ on Σ was bipartite as the mixed states of ϕ on $\Sigma - \Sigma \cap \mathcal{B}$ and $\Sigma \cap \mathcal{B}$. Therefore, the otherwise lost information can be understood to be enclosed in a Black hole remnant. In the next subsections, we will introduce two remnant candidates: Planckons and Cornucopions.

4.5.1 Planckons

The first remnant proposal was presented by Aharonov, Casher & Nussinov in [6]. They suggested that the final state of Black hole evaporation would be a stable[13] remnant whose mass is near Planck mass M_{pl}. They called those objects "planckons". In the proposal they do not discuss further the phenomenology of the remnant or its undelying physics.

The stability of the planckon can be justified by how it would emit its energy if it was to evaporate. The typical energy of a planckon should be of the order of Planck energy E_{pl}.

[13]In this context, stable should be understood as something that lasts longer than the age of the universe.

If the Black hole holds sufficient correlations with previously emitted radiation in order to keep the full state pure, its Von Neumann entropy should be

$$S = 4\pi K_B \left(\frac{M}{M_{pl}}\right)^2. \tag{4.24}$$

Therefore, the amount of information contained in the plackon is

$$I = 4\pi \left(\frac{M}{M_{pl}}\right)^2. \tag{4.25}$$

Assuming that each emitted particle carries a unit of information with it, the number of quanta that must be emitted, up to dimensionless constant factors, can be estimated to be

$$N \approx \left(\frac{M}{M_{pl}}\right)^2, \tag{4.26}$$

and the average energy of each of quantum is

$$\varepsilon = \frac{E_{pl}}{N} = E_{pl} \left(\frac{M_{pl}}{M}\right)^2. \tag{4.27}$$

According to Heisenberg's uncertainty principle, the amount of time τ necessary to emit a quantum of that energy must be such that

$$\tau \geq \frac{\hbar}{2\varepsilon} \sim \tau_{pl} \left(\frac{M}{M_{pl}}\right)^2, \tag{4.28}$$

where τ_{pl} is the Planck time. That quantum should present at most a weak correlation with other remnant quanta, therefore, they should be emitted one at a time to avoid overlapping between their wave-functions. With that, one gets a remnant lifetime of at least

$$\tau_r \sim \tau_{pl} \left(\frac{M}{M_{pl}} \right)^4 , \tag{4.29}$$

where $\tau_r = N\tau$. As the mass M of the initial Black hole could have been arbitrarily large, one can treat remnants as stable objects. It is important to remark that this estimation was originally presented in [106] and it is the most optimistic one, i.e., it is the one that predicts the lowest lifetime. There are at least three other, less optimistic, estimations: One by Hawking [107], one by Giddings [108] and another by Aharonov, Casher & Nussinov [6]. Therefore, if a planckon exists, it should have a huge lifetime.

4.5.2 Cornucopions

The number of internal states of a planckon can be estimated, using Boltzmann entropy, as

$$N_{int} = e^{\frac{S}{k_B}} = e^{4\pi \left(\frac{M}{M_{pl}} \right)^2} , \tag{4.30}$$

which results in $N_{int} \sim 10^{17}$ states for an initial Black hole of one kilogram. What is impressive is not the huge amount of internal states by itself, but the fact that, according to a distant observer, they are necessary to describe an object of planck-volume. A planckon formed by a kilogram Black hole has amazing 10^{17} possible configurations while occupying a volume of approximately[14] $10^{-105} m^3$. That massive spatial density of states is a significant drawback for the remnant hypothesis.

The Hawking temperature of a Reissner-Nordstrom Black hole, in $k_B = \hbar = G = c =$

[14] This is just the volume of a sphere with the radius equal to that of a kilogram mass Schwarzschild Black hole.

1 units, is given by

$$T_H = \frac{\sqrt{M^2 - Q^2}}{2\pi\left(M + \sqrt{M^2 - Q^2}\right)^2}, \tag{4.31}$$

where Q is the charge. Note that its Hawking temperature smoothly tends to zero as $M \to Q$. As a Black hole emits Hawking radiation it would naturally approach such limit because the mass decreases due to radiation emission while the electric charge of a Black hole is expected to be very small, therefore, at some moment it will reach the limit $M \to Q$. That is an indication that an extremal Reissner-Nordstrom Black hole is a remnant candidate, because when it happens its temperature will tend to zero and the evaporation shall stop. This idea was presented in [109].

Unfortunately, analyzing the back-reaction problem in a Reissner-Nordstrom background geometry stills a challenging problem to this day. On the other hand, a striking similarity between a massless scalar field collapse in 4-dimensions and dilaton-gravity coupled to matter in 2-dimensions was noted in [109]. Such 2-dimensional theory is much more tractable and was explored in [110]. That is the so-called CGHS model, whose action is[111]

$$S = \frac{1}{2\pi}\int \mathrm{d}^2x\,\sqrt{-g}\left\{e^{-2\phi}\frac{1}{G_2}\left[R + 4\left(\boldsymbol{\nabla}\phi\right)^2 + 4\lambda^2\right] - \frac{1}{2}\left(\boldsymbol{\nabla}f\right)^2\right\}, \tag{4.32}$$

where R is Riemann scalar curvature, G_2 is the 2-dimensional gravitational constant, λ is the cosmological constant, f is a massless scalar field and ϕ is a dilaton field. In [110] the authors showed that their model admits a solution that is analogous to a near-extremal Black hole in 4-dimensions but it was not well suited to study the backreaction problem. The expansion parameter for the weak-field regime is e^{ϕ}, where ϕ is the dilation field. In the original CGHS model, after the radiated energy becomes equal to the initial mass of the collapsing field, that expansion parameter is bigger than unity, therefore an expansion around such parameter would diverge. On the other hand, they note that by adding a large number N of matter fields to the theory that expansion parameter would become small, making it possible to use that to explore the backreaction problem. The proposed effective action for N fields CGHS is

$$S_{eff} = \frac{1}{2\pi}\int \mathrm{d}^2x\,\sqrt{-g}\left\{e^{-2\phi}\frac{1}{G_2}\left[R + 4\left(\boldsymbol{\nabla}\phi\right)^2 + 4\lambda^2\right] - \frac{1}{2}\left(\boldsymbol{\nabla}f\right)^2\right\} + NS_P, \tag{4.33}$$

$$S_P = -\frac{\hbar}{96\pi}\int \mathrm{d}^2x\,\sqrt{-g}R\,\Box^{-1}R \tag{4.34}$$

where S_P is the Polyakov effective action, i.e., the action that results in the stress-energy tensor that can be derived from the trace anomaly[70].

This extended model was explored in [112], where the authors coined the term cornucopion to refer to the structure of the remnant speculated to exist in the many field CGHS model. That structure would be composed of a flat spacetime connected to another asymptotic region by an infinite throat, as represented in figure 4.11. Such structure would be capable of holding a huge amount of information while occupying a very small region of spacetime according to a distant observer. The infinite throat would be a repository for the infinite degrees of freedom that are necessary to store all information otherwise thought to be lost. For an observer in the asymptotically flat region, the cornucopion would look like a pointlike object. Unfortunately they concluded that the many field CGHS model does not avoid the formation of a singularity, and consequently, the problem of information loss in that singularity.

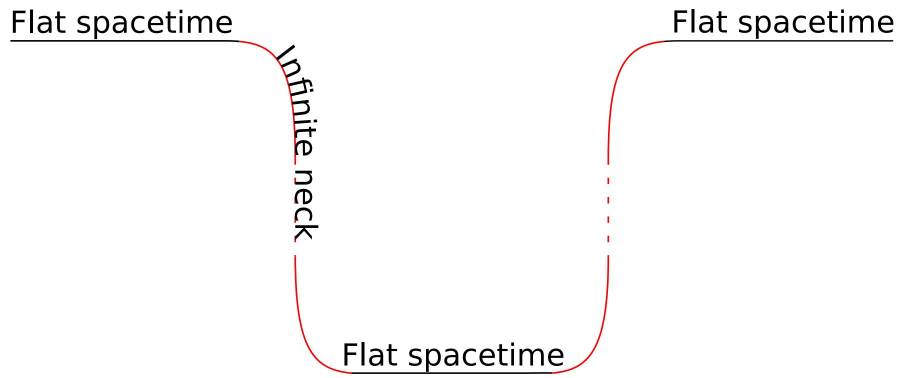

Figure 4.11: Pictoric representation of a cross section of a spacetime contaning a cornucopion. The black lines represent flat assymptotic regions and the red lines represent the infinite neck connecting them.

The arguments against remnants are:

- There is too much information inside a remnant: The Bekenstein-Hawking entropy of a remnant would be very small due to its Planckian mass. If the Bekenstein-Hawking entropy is assumed to be the thermodynamic entropy of a Black hole, it would also be an upper boundary for its Von Neumann entropy. Due to that assumption, only a small number of internal states are to be expected from a remnant which will not be able to store sufficient correlations with Hawking radiation to keep the state of the full system pure.

- Infinite pair production issue: The remnants have small energy and a very large number of internal states in order to hold sufficient correlations with Hawking radiation to keep the state of the full system pure. These two properties would imply that remnants are common objects, in the sense that they would be easily produced in scattering process. This is not the case as such objects have not been detected to this day.

4.6 Information in Hawking radiation

There is a possibility that information leaks out through Hawking radiation. It is argued that backreaction effects might be able to correct the state of radiation making it pure.

Using the CGHS model, Giddings argued that the corrections induced by back reaction effects did not indicate, to leading order, a recovery of information[113]. As a response to these argument, an article [87] was published by Page, where the aim was to show that the perturbative argument used by the authors of [113] against information recovery is not consistent.

The importance of Page's paper is that he showed that information can possibly leak so slowly that a perturbative approach would not be able to recover it. Moreover, it motivated the usage of Qubit models to understand Black hole evaporation as will be explained in the next chapter.

5

Qubits and Black Holes

There are other possibilities in order to study the black hole information puzzle. A very popular tool in quantum information is the qubit, i.e., a two-level system. It is a natural system to deal with quantum information due to its binary behavior. By using qubit models it is possible to explore the information flow in a physical system, opening a possibility of understanding at least how information can possibly leak out of Black holes. In this chapter we will present a motivation for the usage of Qubit models in the context of black hole evaporation and a first approach to model its thermodynamic behavior using Qubits.

5.1 Page's argument

In 1992, Giddings & Nelson used the CGHS model to argue that information about the initial state of the matter that formed a black hole is not recovered by considering backreaction [113]. In response, Page showed that there is a loophole in the perturbative approach used by Giddings & Nelson [87]. Page's reasoning resulted in a description of black hole evaporation that is more familiar for Quantum Information theorists.

5.1.1 The model

Page presented a phenomenological model of black hole evaporation where the information is recovered. His argument was that the information might leak out so slowly that a perturbative approach would not be able to detect it. In a more precise way, he showed that within his model, the information outflow as a function of time is not analytic in time which implies that a perturbative analysis would conclude that the information outflow is zero, i.e., the information outflow Page is referring to is $\propto\ e^{-\frac{4\pi}{y^2}}$. That function is not analytic at $y = 0$ and therefore, a perturbative analysis would never recover it.

The way Page treated Black hole evaporation motivated a lot of qubit models for black hole evaporation. In this section Page's model will be presented using a Schwarzschild Black hole.

To begin with, we assume that the Bekenstein-Hawking entropy is the thermodynamic entropy of a Black hole. Using the expression 2.22,

$$S = K_B \ln \Omega,$$

for the thermodynamic entropy, the number n of accessible micro-states can be estimated to be

$$n = e^{\frac{S}{K_B}}, \tag{5.1}$$

where S is the Boltzmann entropy of the system. Then, using the equation 3.28,

$$S_{BH} = k_B \frac{c^3}{4\hbar G} A,$$

together with $A = 4\pi R_s^2$, where $R_s = \dfrac{2GM}{c^2}$ is the Schwarzschild radius, we conclude that a Black hole of mass M should have

$$n\left(M\right) = e^{4\pi\left(\frac{M}{M_{pl}}\right)^2} \tag{5.2}$$

accessible internal states. For amusement of the reader, a solar-mass, $M_\odot \sim 10^{30} kg$, black hole would have $n \sim e^{10^{77}}$ accessible micro-states in an area[1] of $A \sim 10^7\ m^2$.

Now consider a Schwarzschild black hole of initial mass M_0. According to Page [114], the mass of that Black hole can be written as a function of time as

$$M\left(t\right) = \left(M_0^3 - 3\frac{\hbar c^4}{G^2}\alpha t\right)^{\frac{1}{3}}, \tag{5.3}$$

[1]According to a distant observer.

where t is the time[2] and α is a coefficient associated with which particle species can be emitted at a significant rate. For simplicity we will take $\alpha = 1$, which does not change the qualitative behavior of the mass function and units $\hbar = c = G = k_B = 1$ for the rest of this section, unless stated otherwise. A plot of the mass of that Black hole as a function of time is shown in figure 5.1.

To continue the analysis we will use Page's conjectured result [115] on the average information in a subsystem. It is as follows: Let a system be such that its Hilbert space has dimension mn, where $m, n \in \mathbb{N}_{>0}$. Then consider a partition in two subsystems represented by two Hilbert spaces of dimensions m and n. Moreover, let the complete system be in a random pure state. In this setting, the average information[3] contained in the subsystem of dimension m is given by

$$I\left(m,n\right) = \begin{cases} \ln m + \dfrac{m-1}{2n} - \displaystyle\sum_{k=n+1}^{mn} \dfrac{1}{k} & \text{, if } n \geq m \\ \ln m + \dfrac{n-1}{2m} - \displaystyle\sum_{k=m+1}^{mn} \dfrac{1}{k} \sim \ln m - \ln n + \dfrac{m}{2n} & \text{, if } n \leq m \end{cases} \quad , \tag{5.4}$$

where the asymptotic expression can be used whenever $m \gg 1$. That result was proven shortly after Page conjectured it and the proofs can be found in [116, 117, 118, 119].

We proceed to model the black hole and radiation as two systems with n and m microstates, respectively. For the black hole microstates, we use equation (5.2) with the mass profile given by expression (5.3) to get $n\left(t\right) = e^{4\pi M(t)^2}$ with $n\left(t\right) = e^{4\pi M(t)^2}$. The number of radiation microstates is then[4] $m\left(t\right) = e^{4\pi\left(M(0)^2 - M(t)^2\right)}$. By plugging that in equa-

[2]According to an observer at $\mathscr{I}^+$.

[3]The average is over all pure joint states of Black hole and radiation.

[4]Under the assumption that the dimension of the joint Hilbert space is constant and equal to $n\left(0\right)$,

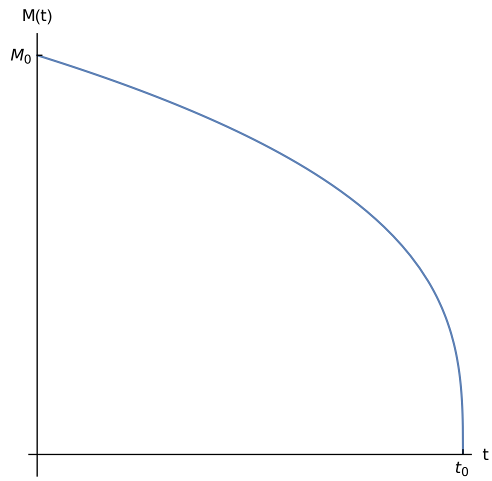

Figure 5.1: Black Hole mass as a function of time according to equation (5.3)

tion (5.4), we get the average information contained in the radiation system as a function of time. With that result one can also calculate the average entanglement entropy as a function of time as

$$S(t) = \ln m(t) - I(t)\,. \tag{5.5}$$

Furthermore, one can estimate the thermodynamic entropy of Hawking radiation to be

$$S_r(t) = \ln m(t)\,. \tag{5.6}$$

The Page time t_{page} in this situation is the turning point of entanglement entropy as a function of time, because at that instant, the entanglement entropy is equal to the Bekenstein-Hawking entropy. In the present model, it is the instant of time when the thermodynamic entropy of radiation is equal to the Bekenstein-Hawking entropy of the black hole. Page time is commonly used as an estimate for the instant of time when information begin to

therefore, $m(t)\,n(t) = n(0)$ which implies $m(t) = \dfrac{n(0)}{n(t)}$.

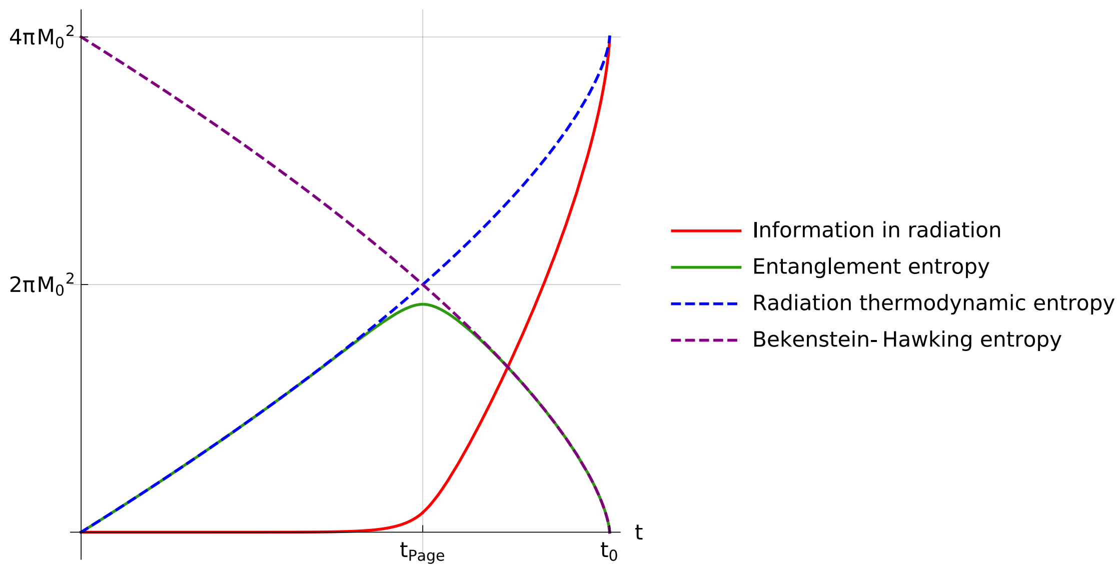

Figure 5.2: In this plot are: Information in radiation given by equation (5.4), the entanglement entropy given by equation (5.5),the radiation thermodynamic entropy given by equation (5.6) and the Bekenstein-Hawking entropy given by (3.28). To estimate the number of accessible micro-states for the Black hole system, equation (5.2) was used with the mass profile given by equation (5.3).

leak from a Black hole.

Putting it all together one can plot the so-called Page curve, which shows a possible behavior for the entanglement entropy between a Black hole and Hawking radiation. A plot of the Page curve for the model in question is presented in figure 5.2. The qualitative behavior of all these quantities is the same for any mass profile used, given that it goes to zero at some instant of time t_0. In figure 5.2, from $t = 0$ up to t_{page}, the Black hole is getting entangled with the Hawking radiation it is emitting, increasing the entanglement entropy. For $t > t_{page}$ the Black hole keeps evaporating, but instead of developing more entanglement, it starts to emit radiation that is correlated to the early radiation. Because of that, the black hole loses the correlations it has with early radiation and the entanglement entropy diminishes. While the Black hole is evaporating, it dwindles which implies a reduction of its

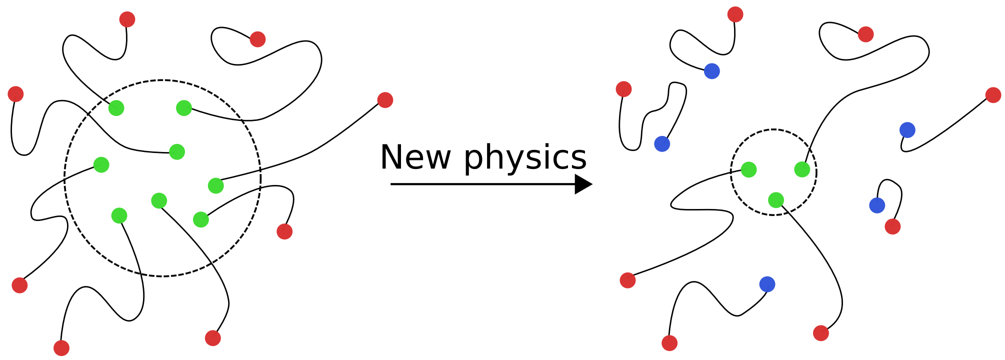

Figure 5.3: In this figure the dashed line is the event horizon, the green dots are pictorial representations Black hole degrees of freedom, the red dots are early radiation quanta, the blue dots are late radiation quanta and the solid lines represent correlation. In Page's scenario, late radiation carries correlations that the Black hole has with early radiation into the radiation system. As effect, all information is brought back to the outside of the Black hole due to evaporation, resulting in no information loss.

Bekenstein-Hawking entropy. At the same time there is energy leaking from the Black hole into the environment in the form of thermal radiation and this is why the thermodynamic entropy of the radiation system is expected to increase during the evaporation process. An illustration of the "disentanglement" process is presented in figure 5.3.

The information loss scenario happens in the semi-classical approach because the entanglement entropy is predicted to be monotonically increasing, therefore the correlations between the Black hole and the radiation are never transferred to the radiation, i.e., it never emits radiation that is correlated with early Hawking radiation[5]. At the same time, the Bekenstein-Hawking entropy is monotonically decreasing, hence, the entanglement entropy will become bigger than the thermodynamic entropy. This is not reasonable as the thermodynamic entropy, here supposed to be equal to the Bekenstein-Hawking entropy, represents an upper bound for the Von Neumann entropy of a given macrostate.

[5]One should understand, in this context, early radiation as being any radiation that was emitted before t_{page}.

5.2 Information flow in a simplified Jaynes-Cummings model

In this section, a model, based upon the Jaynes-Cummings model, will be presented and we will show that its behavior is similar to what is expected for an evaporating Black hole. We emphasize that this is not a model of a Black hole but rather a system with similar behavior in terms of the Page curve presented in figure 5.2. The Jaynes-Cummings model is commonly used in quantum optics to study the interaction of atoms with external fields. For more details on that model, the reader is referred to [120] and [121].

We start with a qubit, a generic two-level system[6]. The dynamics of a two-level system can be described by the following Hamiltonian:

$$H_q = \hbar\omega^a\sigma_z, \tag{5.7}$$

where ω^a is the frequency of the qubit and σ_z is the Pauli z-matrix. The eigenstates of that Hamiltonian are the Pauli z-matrix eigenvectors $|0\rangle$ and $|1\rangle$, as explained in appendix B.1. The Qubit eigenstates can then be written as

$$|k\rangle = (\sigma_+)^k\,|0\rangle\,, \tag{5.8}$$

where $k \in \{0, 1\}$ and σ_+ is the Pauli creation operator. The Hilbert space spanned by those eigenstates is $\mathcal{H}_q = \mathrm{span}\,\{|0\rangle\,, |1\rangle\}$.

Not considering interactions, generalizing the qubit Hamiltonian to the case of N qubits

[6] There are several quantum systems that can be used as a qubit, a few examples are: A finite subset of Quantum Harmonic Oscillator states, the number of photons in an electromagnetic cavity and atomic spins. For more information on experimental realizations, the reader is referred to [10].

is straightforward. The N-qubits Hamiltonian is

$$H_q^N = \sum_{i=1}^{N} \hbar\omega_i^a \sigma_i^z, \tag{5.9}$$

where σ_i^z and ω_i^a are respectively the Pauli z-matrix acting only on the i'th qubit and the frequency of the i'th qubit. Its eigenstates can be written as

$$|k_1, \ldots, k_N\rangle = \bigotimes_{i=1}^{N} \left(\sigma_+^i\right)^{k_i} |0\rangle \,, \tag{5.10}$$

where $k_i \in \{0, 1\}$. The Hilbert space spanned by those eigenstates is $\mathcal{H}_q^N = \bigotimes_{i=1}^{N} \mathcal{H}_q$.

The same kind of construction can be used to model M Quantum Harmonic Oscillators. The Hamiltonian for that system is

$$H_F = \sum_{i=1}^{M} \hbar\omega_i^f \left(a_i{}^\dagger a_i + \frac{1}{2}\right), \tag{5.11}$$

where ω_i^f,$a_i{}^\dagger$ and a_i are, respectively, the frequency, the creation operator and the annihilation operator associated to the i'th oscillator. Its eigenstates are

$$|x_1, \ldots, x_M\rangle = \bigotimes_{i=1}^{M} \frac{\left(a_i{}^\dagger\right)^{x_i}}{\sqrt{x_i!}} |0\rangle \,, \tag{5.12}$$

where $x_i \in \mathbb{N}$ and $|0\rangle$ is the eigenstate of the number operator $N_i = a_i^\dagger a_i$ with eigenvalue 0. The Hilbert space spanned by those eigenstates is $\mathcal{H}_F = \bigotimes_{i=1}^{M} \operatorname{span}\{|j\rangle\}_{j=0}^{\infty}$.

One can straightforwardly write the hamiltonian for a system of N qubits and M oscillators because H_F and H_q^N are defined on different spaces, therefore they commute with one another and can be simultaneously diagonalized. As there are no interactions, the eigenstates of $H_F + H_q^N$ is a juxtaposition of eigenstates of H_F and H_q^N:

$$|k_1, \dots, k_N; x_1, \dots, x_M\rangle = \left(\bigotimes_{i=1}^{N} \left(\sigma_+^i\right)^{k_i} |0\rangle \right) \otimes \left(\bigotimes_{i=1}^{M} \frac{\left(a_i{}^\dagger\right)^{x_i}}{\sqrt{x_i!}} |0\rangle \right), \tag{5.13}$$

where $k_i \in \{0, 1\}$ and $x_i \in \mathbb{N}$.

To add some non-trivial dynamics, we use the following interaction Hamiltonian:

$$H_{int} = \sum_{i=1}^{N} \sum_{j=1}^{M} \alpha_{ij} \left(\sigma_+^i a_j + \sigma_-^i a_j{}^\dagger\right), \tag{5.14}$$

where σ_-^i and σ_+^i are, respectively, the Pauli annihilation and creation operators acting on the i'th qubit, and α_{ij} are the coupling constants between oscillator modes and qubits. As there is no explicit time dependence, the total energy of the system is conserved. Moreover, it is easy to see that

$$\left[H_q^N + H_F, H_{int}\right] = 0, \text{ if } \omega_i^a = \frac{\omega_i^f}{2}. \tag{5.15}$$

For simplicity, from now on we will set $\omega_i^a = \dfrac{\omega_i^f}{2}$. As a consequence of that commutation relation, both the interaction and free Hamiltonians can be diagonalized by the same set of eigenstates, pertaining to the Hilbert space $\mathcal{H} = \mathcal{H}_q^N \otimes \mathcal{H}_F$.

The time evolution operator, which relates the state of the system at time t_0 to that at time t, is

$$U(t, t_0) = e^{-i\frac{H_F + H_q^N + H_{int}}{\hbar}(t - t_0)}. \tag{5.16}$$

As a consequence of (5.15), one can write is as

$$U\left(t,t_0\right)=e^{-i\frac{H_F+H_q^N}{\hbar}(t-t_0)}e^{-i\frac{H_{int}}{\hbar}(t-t_0)}. \tag{5.17}$$

Then, given a state $|\psi\rangle \in \mathcal{H}$ one can say that after an interval of time $t-t_0$ the state of the system will be given by

$$|\psi\left(t,t_0\right)\rangle = U_{free}\left(t,t_0\right)U_{int}\left(t,t_0\right)|\psi\rangle\,, \tag{5.18}$$

where

$$U_{free}\left(t,t_0\right)=e^{-i\frac{H_F+H_q^N}{\hbar}(t-t_0)} \tag{5.19}$$

and

$$U_{int}\left(t,t_0\right)=e^{-i\frac{H_{int}}{\hbar}(t-t_0)}. \tag{5.20}$$

One can then define interaction-picture states as

$$|\psi\left(t,t_0\right)\rangle_I = U_{free}\left(t,t_0\right)^{-1}|\psi\left(t,t_0\right)\rangle = U_{int}\left(t,t_0\right)|\psi\rangle\,. \tag{5.21}$$

The time evolution of these states is described by the interaction Hamiltonian.

As H_{int} is an Hermitian operator, it is always possible to write it as

$$H_{int}=T^{-1}H_{Dint}T, \tag{5.22}$$

where H_{Dint} is the diagonalized H_{int} and T is a unitary matrix. Therefore, given a quantum state $|\phi\rangle \in \mathcal{H}$ one can write it on the diagonal basis as $|\phi\rangle_D = T|\phi\rangle$. The time evolution operator on this basis is

$$U_{int}^{D}\left(t,t_0\right)=e^{-i\frac{H_{Dint}}{\hbar}\left(t-t_0\right)}. \tag{5.23}$$

Hence, given any quantum state $\left|\phi\right\rangle \in \mathcal{H}$, the time-evolved state is given, in the original basis, by

$$\left|\phi\left(t\right)\right\rangle = T^{\dagger}U_{int}^{D}\left(t\right)T\left|\phi\right\rangle . \tag{5.24}$$

The time evolved density operator can then be calculated as

$$\rho\left(t\right)=T^{\dagger}U_{int}^{D}\left(t\right)T\rho_0T^{\dagger}U_{int}^{D}\left(t\right)^{\dagger}T, \tag{5.25}$$

where $\rho_0=|\phi\rangle\langle\phi|$ and for simplicity $t_0=0$.

By taking a partial trace over the time dependent density operators of the full system, one can calculate the time-dependent reduced density operators of the qubits and oscillators as

$$\rho_q\left(t\right)=\mathrm{Tr}_{\mathcal{H}_F}\left[\rho\left(t\right)\right] \tag{5.26}$$

$$\rho_F\left(t\right)=\mathrm{Tr}_{\mathcal{H}_q^N}\left[\rho\left(t\right)\right]. \tag{5.27}$$

Moreover, one can use that result to calculate the average energy of both systems as a function of time:

$$E_q\left(t\right)=\mathrm{Tr}\left(H_q^N\rho_q\left(t\right)\right), \tag{5.28}$$

$$E_f\left(t\right)=\mathrm{Tr}\left(H_F\rho_F\left(t\right)\right), \tag{5.29}$$

and since the total energy of the system is conserved, it is true that $E_q(t) + E_f(t) = E_0$, where E_0 is the energy of the full system. It is also possible to calculate their entanglement entropy as a function of time:

$$S_e(t) = -\operatorname{Tr}\left[\rho_q(t)\ln\left(\rho_q(t)\right)\right] = -\operatorname{Tr}\left[\rho_F(t)\ln\left(\rho_F(t)\right)\right]. \tag{5.30}$$

To evaluate the Page curve, we should subtract the entanglement entropies from the thermodynamic entropies of both systems. Then, it is necessary to calculate the thermodynamic entropies of the qubits and the oscillators as a function of time. We first have to evaluate the partition functions of both systems. To do that, we will use the canonical ensemble with the time-dependent energies from equations (5.28) and (5.29). Note that after tracing out one of the systems, for instance, the oscillators, we are treating the remaining system as free qubits in contact with an environment represented by the oscillators. Therefore, one need only to solve the thermodynamics of the free systems.

The partition function of a free qubit of frequency ω^q is

$$Z = e^{-\beta\hbar\omega^q} + e^{\beta\hbar\omega^q} = 2\cosh\left(\beta\hbar\omega^q\right), \tag{5.31}$$

where $\beta = (k_B T)^{-1}$ and T is the temperature of the system. For N non-interacting qubits, the partition function is the product of the individual partition functions,

$$Z_q = 2^N \Pi_{i=1}^N \cosh\left(\beta\hbar\omega_i^q\right). \tag{5.32}$$

The average energy of the system is then given by

$$E_q^T = -\frac{\partial \ln Z_q}{\partial \beta} = -\sum_{i=1}^{N} \hbar\omega_i^q \tanh\left(\beta\hbar\omega_i^q\right). \tag{5.33}$$

Assuming that the frequencies of every qubit is equal to ω^q, the average energy is given by

$$E_q^T = -N\hbar\omega^q \tanh\left(\beta\hbar\omega^q\right). \tag{5.34}$$

Solving for the temperature, we get

$$T = \frac{\hbar\omega^q}{k_B}\frac{1}{\tanh^{-1}\left(-\frac{E_q^T}{N\hbar\omega^q}\right)}. \tag{5.35}$$

The thermodynamic entropy is given by

$$S_q^T = k_B \ln Z_q + \frac{E_q^T}{T} = \frac{E_q^T}{T} + Nk_B \ln\left[2\cosh\left(\beta\hbar\omega^q\right)\right]. \tag{5.36}$$

By using the expression (5.35) for the temperature as a function of the energy, one can write this thermodynamic entropy as a function of the energy only.

The partition function of a single oscillator of frequency ω^f is given by

$$Z = \sum_{n=0}^{\infty} e^{-\beta\hbar\omega^f\left(n+\frac{1}{2}\right)} = \frac{1}{2}\,\mathrm{csch}\left(\beta\frac{\hbar\omega^f}{2}\right). \tag{5.37}$$

For M non-interacting oscillators, the partition function is

$$Z_f = \frac{1}{2^M} \Pi_{i=1}^M \operatorname{csch}\left(\beta \frac{\hbar\omega_i^f}{2}\right). \tag{5.38}$$

The average energy in the field from the thermodynamic point of view is given by

$$E_f^T = -\frac{\partial \ln Z_f}{\partial \beta} = \sum_{i=1}^{M} \frac{\hbar\omega_i^f}{2} \coth\left(\beta \frac{\hbar\omega_i^f}{2}\right). \tag{5.39}$$

Assuming that the frequency of every oscillator is equal to ω^f, the energy is given by

$$E_f^T = M \frac{\hbar\omega^f}{2} \coth\left(\beta \frac{\hbar\omega^f}{2}\right). \tag{5.40}$$

Solving for the temperature, one gets

$$T = \frac{\hbar\omega^f}{2k_B} \frac{1}{\tanh^{-1}\left(\frac{M\hbar\omega^f}{2E_f^T}\right)}. \tag{5.41}$$

By plugging equations (5.41) and (5.40) into equation (2.40), we get the following expression for the thermodynamic entropy of the oscillators:

$$S_f^T = k_B \ln Z_f + \frac{E_f^T}{T} = \frac{E_f^T}{T} + Nk_B \ln\left[\frac{1}{2} \operatorname{csch}\left(\beta \frac{\hbar\omega^f}{2}\right)\right]. \tag{5.42}$$

Again, by using the expression (5.41) of the temperature as a function of the energy, it is possible to write this thermodynamic entropy as a function of the energy only.

In order to evaluate the thermodynamic entropy of the time-evolving system, we set the energies of the canonical ensemble to be equal to the time dependent energies of the time-evolving system:

$$E_f^T = E_f(t) \tag{5.43}$$

$$E_q^T = E_q(t)\,. \tag{5.44}$$

With all that set, the major difficulty is to diagonalize the interaction Hamiltonian. One can overcome that difficulty with the help of any software capable of diagonalizing matrices. We used Mathematica to simulate the time evolution of a system of two qubits and three oscillator modes. We have set $\omega^q = \omega^f = 1$ and $\alpha_{i,j} = 1$. With this setup, we evaluated the time evolution of three distinct states $|\psi\rangle_1$, $|\psi\rangle_2$ and $|\psi\rangle_3$, where

$$|\psi\rangle_1 = |1,0\rangle \otimes |0,0,0\rangle\,, \tag{5.45}$$

$$|\psi\rangle_2 = \frac{1}{\sqrt{2}}\left(|1,0\rangle + |0,1\rangle\right) \otimes |0,0,0\rangle\,, \tag{5.46}$$

$$|\psi\rangle_3 = \frac{1}{\sqrt{2}}\left(|1,0\rangle \otimes |0,0,0\rangle + |0,0\rangle \otimes |1,0,0\rangle\right). \tag{5.47}$$

The resulting Page curves are presented in figures 5.4, 5.5 and 5.6. We interpret the qubits as being analogous to the internal states of the Black hole and the oscillators as being analogous to the radiation states. Moreover, in order to mimic an evaporating system, we had to choose states such that the qubits ("Black hole") have non-zero initial energy.

By comparing the page curves we see that, at least for this model and these states, the qualitative behavior of the thermodynamic entropy, entanglement and information do not vary by choosing different states. Moreover, we see that the dynamics of this model is capa-

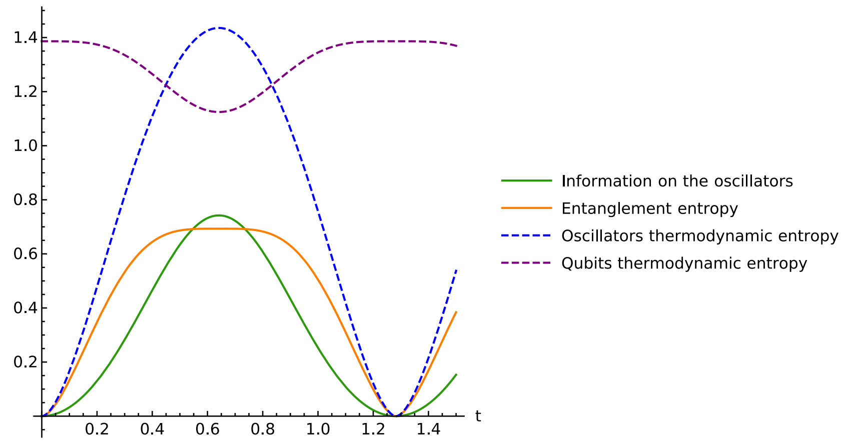

Figure 5.4: Page curve for the initial state $|\psi\rangle_1 = |1,0\rangle \otimes |0,0,0\rangle$.

ble of transferring entanglement between qubits to the oscillators as seen in figures 5.4 and 5.5. In state (5.45) there is no initial entanglement in any system, however the qubit system and the oscillators develop some degree of entanglement until reaching a maximum and then they get separated. In state (5.46) there is explicit initial entanglement between two qubits and again, they develop correlation with the oscillators and after reaching a maximum, they get separated. These two cases do present a similar behavior to that expect in Page's model 5.2. The drawback is that in the Black hole case, we expect early radiation to be entangled with internal black hole states, i.e., there is entanglement between parts of different systems. The state (5.47) has that kind of initial inter-system entanglement but the dynamics of our model was barely able to change its entanglement entropy as it stayed almost constant, as seen in figure 5.6.

While not being an accurate model for black hole evaporation it serves as a proof of principle that it might be possible to follow that line and produce a reliable model with the desired dynamics that would more closely resemble what is expected of a black hole. Moreover, the dynamics we have considered is fairly simple, it consists of letting the qubits inter-

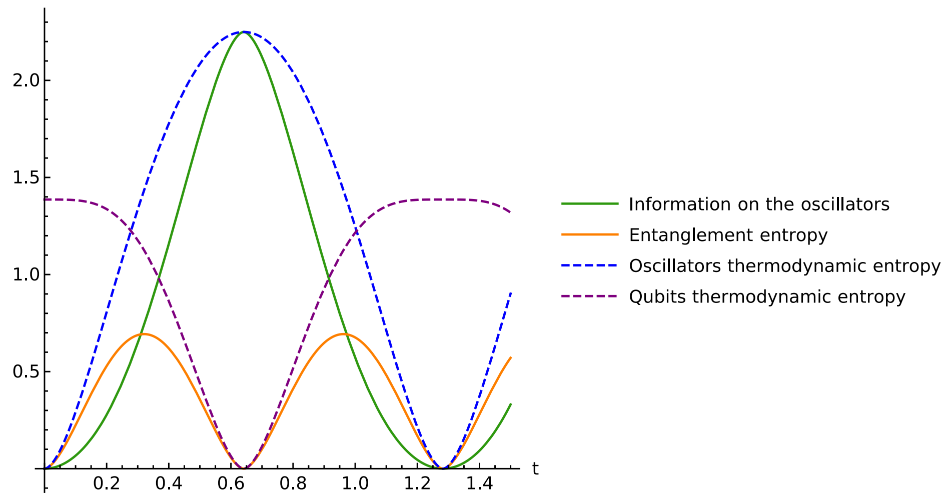

Figure 5.5: Page curve for the initial state $|\psi\rangle_2 = \dfrac{1}{\sqrt{2}} \left(|1,0\rangle + |0,1\rangle\right) \otimes |0,0,0\rangle$.

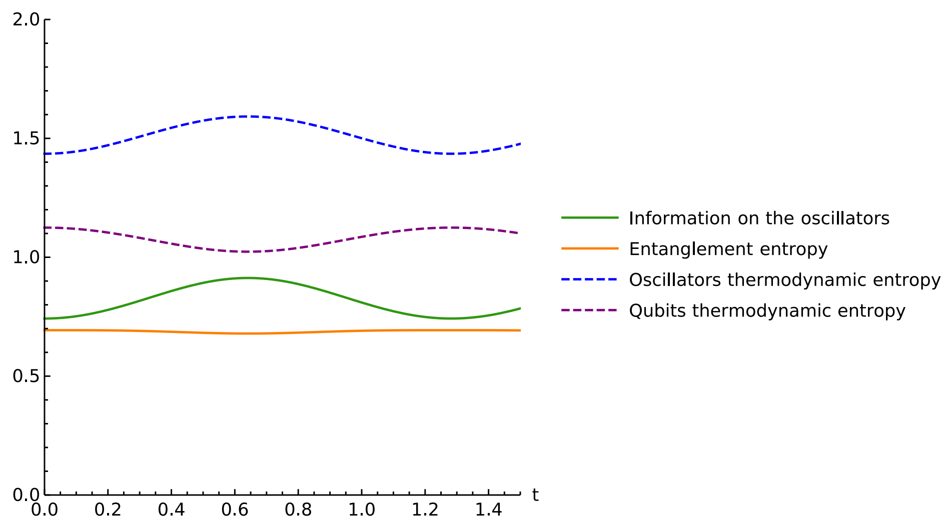

Figure 5.6: Page curve for the initial state $|\psi\rangle_3 = \dfrac{1}{\sqrt{2}} \left(|1,0\rangle \otimes |0,0,0\rangle + |0,0\rangle \otimes |1,0,0\rangle\right)$.

act with any oscillator. A more interesting behavior might arise if, for instance, we let the oscillators interact with themselves. In that scenario we would expect to see some behavior that resembles a “scrambling”, i.e., non-trivial entanglement dynamics, of quantum information in the oscillators.

Regarding a key limitation in our simulation, we were not able to use any state containing more than one energy quanta as it would cause the qubit system to reach negative temperatures and such behavior is not expected in the Hawking temperature of a black hole. The number of accessible microstates for a system of N qubits from which N_e of them are excited is given by

$$\Omega\left(N, N_e\right) = \frac{N!}{N_e!\left(N - N_e\right)!}. \tag{5.48}$$

A diverging temperature happens when

$$\frac{\mathrm{d}\Omega}{\mathrm{d}N_e} = 0, \tag{5.49}$$

and negative temperatures happen whenever

$$\frac{\mathrm{d}\Omega}{\mathrm{d}N_e} < 0, \tag{5.50}$$

The maximum number of energy quanta that can be added to a system of N qubits in order to keep its temperature strictly positive and non-diverging is given by the greatest integer less than or equal to $\frac{N}{2}$, i.e.,

$$N_{max} = \left\lfloor \frac{N}{2} \right\rfloor. \tag{5.51}$$

It is clear that adding more qubits would allow us to simulate more energetic states but it

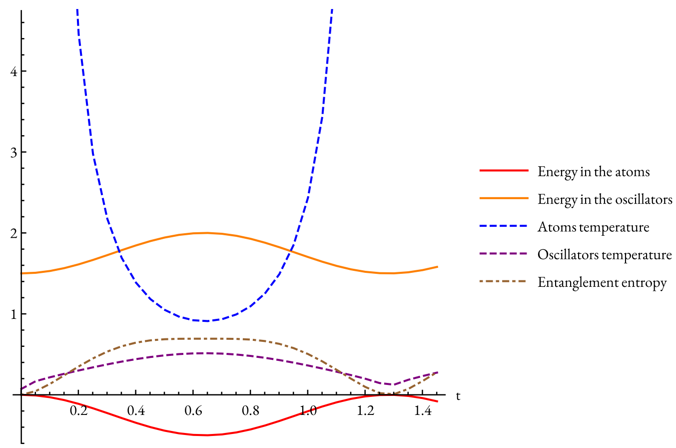

Figure 5.7: Energies, temperatures and entanglement entropy for the initial state $|\psi\rangle_1 = |1, 0\rangle \otimes |0, 0, 0\rangle$.

was not viable given the available computing power we had at this time. Given a more powerful computer it would have been possible to try out configurations with higher energy.

Another interesting characteristic of $|\psi\rangle_1$,$|\psi\rangle_2$,$|\psi\rangle_3$ in this model can be noted by looking at the plots of energy, temperature and entanglement entropy, presented in figures 5.7, 5.8 and 5.9. All these figures present an odd behavior, for an interval of time, where the heat flows in the wrong direction, i.e., the hotter system gets hotter while the colder system gets colder. That behavior is observed only when the entanglement entropy is diminishing, which may indicate that the correlations between the oscillators and qubits are being converted into energy, which provokes that odd behavior in the temperatures. Experimental evidence of the conversion of correlations into energy was presented, although in a different setup, in [122, 123, 124].

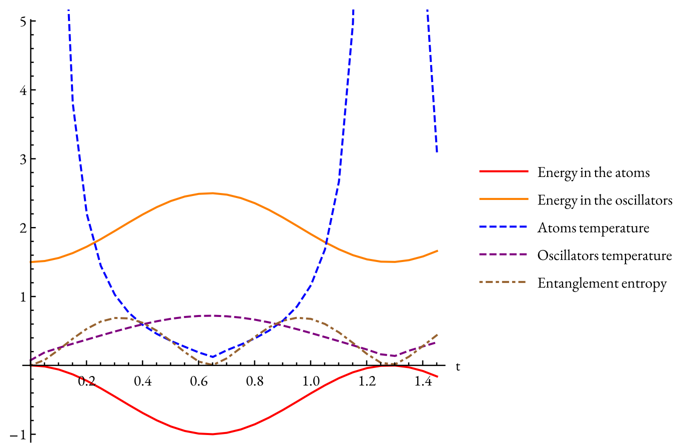

Figure 5.8: Energies, temperatures and entanglement entropy for the initial state $|\psi\rangle_2 = \frac{1}{\sqrt{2}}\left(|1,0\rangle \otimes |0,0,0\rangle + |0,1\rangle \otimes |0,0,0\rangle\right)$.

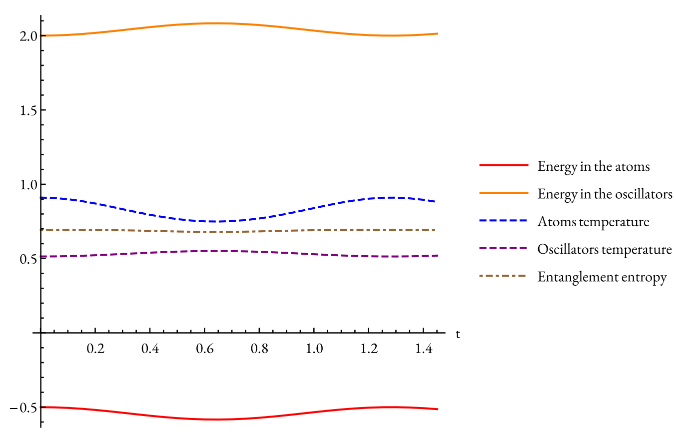

Figure 5.9: Energies, temperatures and entanglement entropy for the initial state $|\psi\rangle_3 = \frac{1}{\sqrt{2}}\left(|1,0\rangle \otimes |0,0,0\rangle + |0,0\rangle \otimes |1,0,0\rangle\right)$.

6

Conclusion

We introduced a brief history together with the minimal knowledge needed to understand the information loss puzzle. That problem is particularly tricky to properly state and even harder to solve, given that there is still no solution up to this day. By writing that text we hope to make that problem more accessible for students engaged in both Quantum Information and General Relativity. There is a high barrier for one that wants to study that topic due to the huge amount of available literature which makes it difficult to follow the discussions and to its interdisciplinary nature, that requires specific knowledge from Statistical mechanics, General Relativity and Quantum Information.

We finished that text with a brief presentation of a Qubit toy model for Black hole evaporation. Due to the simplicity of the presented model, it serves as a proof of concept that the kind of reasoning used might lead to a deeper understanding of the dynamics underlying information transfer in Black hole evaporation. We hope to continue developing the model in order for it to represent more closely the behavior that is expected from an evaporating Black hole.

A

Details of Quantum Mechanics

A.1 Boundaries for the eigenvalues of density operators

The hermiticity of the density operators imply that any eigenvalue, λ, is a real number. Furthermore, the positive semi-definiteness of the density operators imply that $\min \lambda / geq 0$.

The trace of a given $\rho \in \mathcal{M}$ is

$$\operatorname{Tr}\rho = \sum_{i=1}^{\dim(\mathcal{M})} \lambda_i = 1, \tag{A.1}$$

where λ_i are the eigenvalues of ρ. As $\lambda_i \geq 0$, the maximum value for any eigenvalue is 1.

For that reason, any eigenvalue is bounded in the interval $\lambda \in [0, 1]$. .

A.2 Boundaries of the inner product norm on the state space

The norm of a given $\rho \in \mathcal{M}$ is

$$\|\rho\|^2 = \operatorname{Tr}\left(\rho^2\right) = \sum_{i=1}^{\dim(\mathcal{H})} \lambda_i^2, \tag{A.2}$$

where $\lambda_i \in [0, 1]$,as showed in A.1, are the eigenvalues of ρ. For a moment, assume that all eigenvalues are smaller than one. Under this assumption, $\lambda_i^2 < \lambda_i$. Therefore

$$\|\rho\|^2 = \sum_{i=1}^{\dim(\mathcal{H})} \lambda_i^2 < \sum_{i=1}^{\dim(\mathcal{H})} \lambda_i = 1, \tag{A.3}$$

which imply that $\|\rho\| < 1$ if there are no eigenvalues that are equal to one. Conversely, assume that one of the eigenvalues is equal to one. In that case, the trace condition, $\operatorname{Tr}\rho = 1$ imply that all other eigenvalues are equal to zero, therefore, under this assumption,$\|\rho\| = 1$. Combining both cases, one concludes that $\|\rho\| \leq 1$.

For the lower bound, by using the method of Lagrange multipliers it is straightforward to minimize $\|\rho\|$, given $\mathrm{Tr}\,\rho = 1$. There is only one solution,

$$\begin{cases} \lambda_1 = \lambda_2 = \cdots = \lambda_{\dim(H)} = \dfrac{1}{\dim(H)} \\ \min \|\rho\| = \dfrac{1}{\dim(H)}. \end{cases} \tag{A.4}$$

Hence, $\|\rho\| \in \left[\dfrac{1}{\dim(H)}, 1\right]$.

A.3 The diagonal representation of a density operator

The hermiticity of any $\rho \in \mathcal{M}$ implies that it is always possible to find an orthonormal basis for the state space where a given ρ is diagonal. In that basis, one can write

$$\rho = \sum_{i=1}^{\dim(\mathcal{H})} \alpha_i \hat{e}_i, \tag{A.5}$$

where $\{e_i\}_{i=1}^{\dim(H)^2}$ is the orthonormal basis set for $\mathcal{M}$ that diagonalizes ρ.

B

Pauli matrices and Quantum states

B.1 Pauli matrices and state space

Pauli matrices are a set of 2×2, unitary and hermitian complex matrices. One can also understand them as generators of the fundamental representation of $SU(2)$ group[1]. Such set of matrices obey

$$[\sigma_a, \sigma_b] = 2i\epsilon_{abc}\sigma_c, \tag{B.1}$$

where ϵ_{abc} is the Levi-Civita symbol. The matrix representation of Pauli matrices is of no mystery, see for instance [10] for a Quantum Information point of view, or [125] for a Quantum Field Theory point of view. Such representation is given by

$$\mathbf{1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{B.2}$$

The normalized eigenvectors of σ_z are given by

$$|1\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, |0\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \tag{B.3}$$

such that $\sigma_z |1\rangle = |1\rangle$ and $\sigma_z |0\rangle = -|0\rangle$. The space spanned by the eigenvectors of σ_z is $\mathcal{H} = \text{span}\{|0\rangle, |1\rangle\}$ and happens to be a Hilbert space with the usual dot product as inner product.

We now explore the space of linear operators acting over $\mathcal{H}$. To do that, one should first compose a basis to such space, this can be done by combining eigenvectors with Kronecker

[1]The actual generators are given by i times the Pauli matrices.

products,

$$|1\rangle \otimes |1\rangle^\dagger = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, |1\rangle \otimes |0\rangle^\dagger = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, |0\rangle \otimes |1\rangle^\dagger = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, |0\rangle \otimes |0\rangle^\dagger = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \tag{B.4}$$

and then writing the Pauli matrices as linear combinations of these products[2],

$$\mathbf{1} = |0\rangle\langle 0| + |1\rangle\langle 1| \tag{B.5}$$

$$\sigma_x = |1\rangle\langle 0| + |0\rangle\langle 1| \tag{B.6}$$

$$\sigma_y = -i\,|1\rangle\langle 0| + i\,|0\rangle\langle 1| \tag{B.7}$$

$$\sigma_z = |0\rangle\langle 0| - |1\rangle\langle 1|\,. \tag{B.8}$$

This is useful because the action of Pauli matrices x and y over a z eigenstate is easily computed,

$$\sigma_x\,|0\rangle = |1\rangle \tag{B.9}$$

$$\sigma_x\,|1\rangle = |0\rangle \tag{B.10}$$

$$\sigma_y\,|0\rangle = -i\,|1\rangle \tag{B.11}$$

$$\sigma_y\,|1\rangle = i\,|0\rangle\,. \tag{B.12}$$

[2] $|a\rangle\langle b| \equiv |a\rangle \otimes |b\rangle^\dagger$.

A generic unitary hermitian linear operator O acting on $\mathcal{H}$ can be written as

$$O = a\mathbf{1} + b\sigma_x + c\sigma_y + d\sigma_z, \tag{B.13}$$

where $a, b, c, d \in \mathbb{C}$ and $\mathbf{1}$ is the identity. One can determine constraints over a, b, c, d in order to define a proper state space.

The condition $O^\dagger = O$ implies that in order to represent a physical state, the coefficients a, b, c, d must be such that $a, b, c, d \in \mathbb{R}$. Moreover, imposing $\operatorname{Tr} O = 1$ demand that $a = \dfrac{1}{2}$, therefore, the most generic operator obeying both constraints is

$$O = \frac{1}{2}\mathbf{1} + b\sigma_x + c\sigma_y + d\sigma_z. \tag{B.14}$$

Using the matrix form, the characteristic polynomial associated with O is

$$\lambda^2 - \lambda + \frac{1}{4} - \left(b^2 + c^2 + d^2\right) = 0, \tag{B.15}$$

and the eigenvalues are

$$\lambda_\pm = \frac{1}{2} \pm \sqrt{b^2 + c^2 + d^2}. \tag{B.16}$$

Imposing that $\lambda_\pm \geq 0$ imply that $\sqrt{b^2 + c^2 + d^2} \leq \dfrac{1}{2}$. To explore that condition one can define

$$r = \sqrt{b^2 + c^2 + d^2} \tag{B.17}$$

$$b = r \sin(\theta) \cos(\phi) \tag{B.18}$$

$$c = r \sin(\theta) \sin(\phi) \tag{B.19}$$

$$d = r \cos(\phi), \tag{B.20}$$

where $\theta \in [0, \pi]$ and $\phi \in [0, 2\pi[$. Therefore, the state space is completely parametrized as,

$$\rho = \frac{1}{2}\mathbf{1} + r\left[\sin(\theta)\cos(\phi)\,\sigma_x + \sin(\theta)\sin(\phi)\,\sigma_y + \cos(\phi)\,\sigma_z\right], \tag{B.21}$$

where $r \in \left[0, \frac{1}{2}\right]$. This is the state space naturally spanned by the Pauli matrices.

C

Entropy and statistical mechanics

C.1 Maximizing entropy given a set of constraints

Let $\hat{A}_i \in \mathcal{A}$ be a set of observables and A_i be the expected value of $\hat{A}_i$. To maximize the entropy considering the constraints $\mathrm{Tr}\left(\rho \hat{A}_i\right) = A_i$ and $\mathrm{Tr}\,\rho = 1$, where ρ denote the ensemble state, one can use the method of Lagrange multipliers, with Lagrangian given by

$$L[\rho] = \mathrm{Tr}(\rho \log \rho) + \sum_i \lambda_i \left(\mathrm{Tr}\left(\rho \hat{A}_i\right) - A_i\right) + \lambda_0 (\mathrm{Tr}\,\rho - 1)\,. \tag{C.1}$$

The density operator that maximize the Lagrangian is given by

$$\rho = e^{-1-\lambda_0-\sum_i \lambda_i \hat{A}_i}. \tag{C.2}$$

As the density operator have trace equals unity, one can use that fact to implicitly calculate the Lagrange multiplier λ_0 to be

$$e^{1+\lambda_0} = \mathrm{Tr}\left(e^{-\sum_i \lambda_i \hat{A}_i}\right). \tag{C.3}$$

One then defines the partition function as

$$Z = \mathrm{Tr}\left(e^{-\sum_i \lambda_i \hat{A}_i}\right), \tag{C.4}$$

allowing one to write the density operator in a more familiar form,

$$\rho = \frac{e^{-\sum_i \lambda_i \hat{A}_i}}{Z}.. \tag{C.5}$$

The partition function plays a central role in statistical mechanics as it encodes the information about the A_i, it is simple to verify that

$$\langle A_i \rangle = -\frac{\partial \ln Z}{\partial \lambda_i}. \tag{C.6}$$

This relation allows one to recover relations between the observables and the Lagrange multipliers.

Moreover, the statistical entropy of the ensemble density operator is given by

$$\begin{aligned} S[\rho] &= -\operatorname{Tr}(\rho \ln \rho) \\ &= -\operatorname{Tr}\left(\rho \ln e^{-\sum_i \lambda_i \hat{A}_i} - \rho \ln Z\right) \\ &= \ln Z \operatorname{Tr} \rho + \operatorname{Tr}\left(\rho \sum_i \lambda_i \hat{A}_i\right) \qquad \text{(C.7)} \\ &= \ln Z + \sum_i \lambda_i A_i, \qquad \text{(C.8)} \end{aligned}$$

as expected, the statistical entropy is dimensionless but one can multiply it by K_B to make it match the thermodynamic entropy.

D

Short dictionary of entropies

D.1 Boltzmann entropy

Boltzmann's entropy is a very important contribution to the development of early statistical mechanics. It relates the thermodynamic entropy S of a physical system with the number of accessible microstates W for a given macrostate $\mathcal{X} = \{X_i\}_{i=1}^N$ as:

$$S(\mathcal{X}) = k_B \ln W. \tag{D.1}$$

D.2 Shannon entropy

In the book [18], the mathematical theory of communications is introduced and therein is an abstract notion of information, independent of Physics. Therefore, at least in principle, it is disconnected from Quantum Mechanics.

The Shannon entropy is defined for any discrete or continuous probability distribution. In this text we will only refer to discrete probability distributions. For a probability distribution $\mathcal{P} = \{p_i\}_{i=1}^N$, the Shannon entropy is defined as

$$H[\mathcal{P}] = -K \sum_{i=1}^{N} p_i \ln p_i, \tag{D.2}$$

where K is a real positive constant that amounts to a choice of unit of measurement.

The Shannon entropy of a probability distribution $\mathcal{P}$ is a measurement of how uncertain one would be of the outcome of an event that has $\mathcal{P}$ as the probability distribution of its outcomes.

D.3 Von Neumann entropy

Arising in the context of Quantum Statistical Mechanics, the Von Neumann entropy is a function $S : \mathcal{L}(\mathcal{H}) \mapsto \mathcal{R}_{\geq 0}$, where $\mathcal{L}(\mathcal{H})$ is the set of operators that act on $\mathcal{H}$, the Hilbert space corresponding to some quantum system.

For any given density operator $\rho \in \mathcal{L}(\mathcal{H})$ representing the quantum state of a physical system, one defines its Von Neumann entropy as

$$S[\rho] = -\operatorname{Tr} \rho \ln \rho. \tag{D.3}$$

The Von Neumann entropy is defined for any density operator, therefore, it is more general than the entanglement entropy or the thermodynamic entropy. For a complete discussion about the Von Neumann entropy, the reader is referred to [11].

D.4 Von Neumann entropy as entanglement entropy

The concept of entanglement was presented in section 1.5, and to quantify it we presented the Entanglement entropy. From equation (1.30), it is easy to see that it is just a special case of Von Neumann's entropy. More precisely, the entanglement entropy can be understood as the Von Neumann entropy of one part of a quantum system that was divided in two parts.

D.5 Von Neumann entropy as a Thermodynamic entropy

It is usually said that the laws of thermodynamics enables one to define a quantity related to the reversibility of a thermodynamic process, the so-called thermodynamic entropy. A more precise statement however, would be that they allow one to define the variation of thermodynamic entropy[1], as presented in section 2.1.

[1] By using the third law of thermodynamics one can make sense of the absolute value of the thermodynamic entropy but only for pure, perfectly crystalline substances.

When setting up the framework of statistical mechanics in section 2.2, we expose the reasoning in [27] to justify the equivalence between Boltzmann's entropy and the thermodynamic entropy. Therefore, by using Boltzmann's entropy, one is able to make sense of the absolute value of the thermodynamic entropy for substances other than pure, perfectly crystalline ones.

Now we can make a brief pause and note that the Von Neumann entropy of some density operator $\rho \in \mathcal{L}(\mathcal{H})$ in its eigenbasis is

$$S[\rho] = -\sum_{i=1}^{\dim(\mathcal{H})} p_i \ln p_i, \tag{D.4}$$

where the $p_i \in [0,1]$ are the eigenvalues of that density operator. Now let $p_i = p$, where $p = \dfrac{1}{N}$ and $N = \dim(\mathcal{H})$. In that case, the Von Neumann entropy of that density operator is

$$S[\rho] = -\frac{1}{N}\sum_{i=1}^{N} \ln N^{-1} = \ln(N). \tag{D.5}$$

One can interpret N as the number of completely distinguishable pure states that can be present in ρ, therefore, what is being calculated is the logarithm of the number of completely distinguishable quantum states pertaining to $\mathcal{L}(\mathcal{H})$. Then, we conclude that the Von Neumann entropy of an ensemble of equiprobable quantum states agrees with its Boltzmann entropy up to Boltzmann's constant. We remark that this equivalence happens only when the density operator is represnting an ensemble of equiprobable quantum states.

The statistical mechanics framework is built on top of the postulate of equal a priori equilibrium probabilities, which is used to describe systems in thermodynamic equilibrium. Therefore, a density operator representing a quantum system in thermodynamic equilibrium will always be such that its Von Neumann entropy is equivalent to its Thermodynamic entropy. From that reasoning, one can think of the Thermodynamic entropy simply as the Von Neumann entropy of a quantum system that is in thermodynamic equilibrium.

D.6 Bekenstein-Hawking entropy

The Bekenstein-Hawking entropy derives from considerations on general relativity, as presented in section 3.5. For a Schwarzschild Black Hole of surface are A, it is given by

$$S_{BH} = k_B \frac{c^3}{4\hbar\pi G} A. \tag{D.6}$$

Although there is still no consensus, it is common to assume that it is the thermodynamic entropy of a Black Hole. That assumption is backed by Hawking's calculations in [3].

References

[1] S. W. Hawking, Physical Review Letters 26, 1344 (1971).

[2] J. D. Bekenstein, Physical Review D 7, 2333 (1973), arXiv:arXiv:1011.1669v3 .

[3] S. W. HAWKING, Nature 248, 30 (1974).

[4] S. W. Hawking, Physical Review D 14, 2460 (1976), arXiv:1601.00921 .

[5] T. Banks, L. Susskind, and M. E. Peskin, Nuclear Physics, Section B 244, 125 (1984).

[6] Y. Aharonov, A. Casher, and S. Nussinov, Physics Letters B 191, 51 (1987).

[7] L. Susskind, L. Thorlacius, and J. Uglum, Physical Review D 48, 3743 (1993).

[8] R. H. Price and K. S. Thorne, *Black Holes: The Membrane Paradigm*, edited by K. S. Thorne, R. H. Price, and D. A. Macdonald (New Heaven: Yale University Press, 1986) p. 1.

[9] A. Almheiri, D. Marolf, J. Polchinski, and J. Sully, (2012), 10.1007/JHEP02(2013)062, arXiv:1207.3123 .

[10] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2010).

[11] I. Bengtsson and K. Zyczkowski, *Geometry of Quantum States: An Introduction to Quantum Entanglement* (2017).

[12] S. Axler, Computer Engineering (1997), 10.1007/b97662, arXiv:arXiv:1011.1669v3 .

[13] O. Christensen, *Functions, Spaces, and Expansions* (Birkhäuser Boston, Boston, 2010).

[14] R. J. Spreeuw, Foundations of Physics 28, 361 (1998).

[15] M. A. Goldin, D. Francisco, and S. Ledesma, Journal of the Optical Society of America B 27, 779 (2010).

[16] X. Song, Y. Sun, P. Li, H. Qin, and X. Zhang, Scientific Reports 5, 1 (2015).

[17] X.-F. Qian, B. Little, J. C. Howell, and J. H. Eberly, Optica 2, 611 (2015), arXiv:1506.01305 .

[18] C. E. Shannon, ACM SIGMOBILE Mobile Computing and Communications Review (2001), 10.1002/j.1538-7305.1948.tb01338.x, arXiv:9411012 [chao-dyn] .

[19] S. Carnot, *Recherche* (1824).

[20] R. S. of Edinburgh., *Transactions of the Royal Society of Edinburgh.*, Vol. v.20 (1850-1853) (Royal Society of Edinburgh., Edinburgh :, 1850) pp. 1–728.

[21] R. Clausius, Annalen der Physik und Chemie 155, 500 (1850).

[22] M. Bailyn, *A survey of thermodynamics* (AIP Press, New York, 1994) p. 657.

[23] R. Fowler, *Statistical thermodynamics : a version of statistical mechanics for students of physics and chemistry.* (Univ. Press, Cambridge, 1939).

[24] C. Carathéodory, Mathematische Annalen 67, 355 (1909).

[25] H. Quevedo, Journal of Mathematical Physics 48, 013506 (2007).

[26] H. B. Callen, *Thermodynamics and an introduction to thermostatistics* (Wiley, 1985) p. 493.

[27] M. Kardar, *Physics Today*, Vol. 62 (2009) pp. 55–56, arXiv:arXiv:1011.1669v3 .

[28] McGraw-Hill, *Engineering*, 2nd ed. (McGraw-Hill Professional, 2003) p. 643.

[29] P. Atkins, *The Laws of Thermodynamics* (Oxford University Press, 2010).

[30] J. C. Maxwell, The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science 19, 19 (1860).

[31] J. C. Maxwell, The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science 20, 21 (1860).

[32] J. W. GIBBS, Proceedings of the American Association for the Advancement of Science , 57 (1884).

[33] J. W. Gibbs, *Elementary Principles in Statistical Mechanics: Developed with Especial Reference to the Rational Foundation of Thermodynamics* (2010).

[34] R. C. Tolman, *The principles of statistical mechanics* (Dover Publications, New York, 1979) p. 660.

[35] R. K. Pathria and P. D. Beale, *Statistical mechanics.* (Academic Press, 2011).

[36] E. M. Purcell and R. V. Pound, Physical Review 81, 279 (1951).

[37] N. F. Ramsey, Physical Review 103, 20 (1956).

[38] M. J. Klein, Physical Review 104, 589 (1956).

[39] I. Bloch, T. Rom, S. Braun, U. Schneider, J. P. Ronzheimer, S. S. Hodgman, and M. Schreiber, Science 339, 52 (2013).

[40] J. Dunkel and S. Hilbert, Nature Physics 10, 67 (2013).

[41] R. H. Swendsen, Reports on Progress in Physics 81, 072001 (2018).

[42] J. A. Beattie and I. Oppenheim, *Principles of thermodynamics* (Elsevier Scientific Pub. Co, Amsterdam ;;New York, 1979) p. 329.

[43] A. B. Pippard, *Elements of Classical Thermodynamics:For Advanced Students of Physics* (Cambridge University Press, 1966) p. 165.

[44] D. J. Evans, E. G. D. Cohen, and G. P. Morriss, Physical Review Letters 71, 2401 (1993).

[45] D. J. Evans and D. J. Searles, Physical Review E 50, 1645 (1994).

[46] D. J. Evans and D. J. Searles, Advances in Physics 51, 1529 (2002).

[47] G. M. Wang, E. M. Sevick, E. Mittag, D. J. Searles, and D. J. Evans, Physical Review Letters 89, 1 (2002).

[48] J. Gieseler, R. Quidant, C. Dellago, and L. Novotny, Nature nanotechnology 9, 358 (2014).

[49] B. Gaveau and L. S. Schulman, Journal of Statistical Physics 145, 1458 (2011).

[50] A. Ramírez-Hernández, H. Larralde, and F. Leyvraz, Physical Review Letters 100, 1 (2008).

[51] A. Ramírez-Hernández, H. Larralde, and F. Leyvraz, Physical Review E - Statistical, Nonlinear, and Soft Matter Physics 78, 1 (2008).

[52] M. N. Bera, A. Riera, M. Lewenstein, and A. Winter, Nature Communications 8, 2180 (2017).

[53] C. G. Knott, *Life and scientific work of Peter Guthrie Tait, supplementing the two volumes of scientific papers published in 1898 and 1900* (Cambridge Univserity Press, 1911).

[54] H. S. Leff and A. F. Rex, *Maxwell's demon : entropy, information, computing* (Princenton University Press) p. 362.

[55] R. Alicki, (2014), arXiv:1406.5879 .

[56] R. Alicki, , 1 (2014), arXiv:1402.2414 .

[57] R. Alicki, , 1 (2018), arXiv:1801.05237 .

[58] R. Alicki and M. Horodecki, Journal of Physics A: Mathematical and Theoretical 52 (2019), 10.1088/1751-8121/ab076f, arXiv:1905.11057 .

[59] S. W. Hawking, Communications in Mathematical Physics 25, 152 (1972), arXiv:1902.09199 .

[60] M. Nakahara, *Text* (2003).

[61] J. Baez and J. P. Muniain, *Gauge Fields, Knots and Gravity*, Series on Knots and Everything, Vol. 4 (WORLD SCIENTIFIC, 1994).

[62] C. W. Misner, K. S. Thorne, J. A. Wheeler, and D. Kaiser, *Gravitation*, p. 1279.

[63] R. M. Wald, *General Relativity.* (University of Chicago Press, 1984) p. 507.

[64] R. Penrose, General Relativity and Gravitation 43, 901 (2011).

[65] B. Carter, Physical Review 141, 1242 (1966).

[66] E. Poisson, *A Relativist's Toolkit* (Cambridge University Press, Cambridge, 2004) p. 190.

[67] E. Gourgoulhon, *Geometry and physics of black holes*, April 2016 (2017).

[68] J. T. Jebsen, General Relativity and Gravitation 37, 2253 (2005).

[69] G. D. Birkhoff, *Relativity and Modern Physics* (Harvard Univserity Press, 1927) p. 283.

[70] A. Fabbri and J. Navarro-Salas, *Modeling black hole evaporation* (2005) p. 334.

[71] S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space–Time* (Cambridge University Press, Cambridge, 1973).

[72] R. Penrose, *Techniques in Differential Topology in Relativity* (Society for Industrial and Applied Mathematics, 1972).

[73] E. Gourgoulhon and J. L. Jaramillo, New Astronomy Reviews 51, 791 (2008).

[74] E. Curiel, Nature Astronomy 3, 27 (2019).

[75] J. M. Bardeen, B. Carter, and S. W. Hawking, Communications in Mathematical Physics 31, 161 (1973).

[76] R. B. Mann, *Black Holes: Thermodynamics, Information, and Firewalls*, SpringerBriefs in Physics (Springer International Publishing, Cham, 2015) pp. 1–95.

[77] R. M. Wald, Living Reviews in Relativity 4, 6 (2001).

[78] M. Heusler, *Black Hole Uniqueness Theorems* (Cambridge University Press, Cambridge, 1996).

[79] A. C. Wall, , 1 (2018), arXiv:1804.10610 .

[80] E. T. Newman, E. Couch, K. Chinnapared, A. Exton, A. Prakash, and R. Torrence, Journal of Mathematical Physics 6, 918 (1965).

[81] E. T. Newman and A. I. Janis, Journal of Mathematical Physics 6, 915 (1965).

[82] P. T. Chruściel, J. L. Costa, and M. Heusler, Living Reviews in Relativity 15, 7 (2012).

[83] R. M. Wald, Directions in General Relativity , 358 (1993), arXiv:9305022 [gr-qc] .

[84] R. Arnowitt, S. Deser, and C. W. Misner, Physical Review 117, 1595 (1960).

[85] D. Wallace, Studies in History and Philosophy of Science Part B - Studies in History and Philosophy of Modern Physics , 1 (2019), arXiv:arXiv:1710.02725v2 .

[86] S. W. Hawking, Communications in Mathematical Physics 46, 206 (1976).

[87] D. N. Page, Physical Review Letters 71, 3743 (1993), arXiv:9306083 [hep-th] .

[88] J. D. Bekenstein, Physical Review D 23, 287 (1981).

[89] R. Bousso, Journal of High Energy Physics 1999, 004 (1999), arXiv:9905177 [hep-th] .

[90] J. D. Bekenstein, Foundations of Physics 35, 1805 (2005).

[91] R. Bousso, Reviews of Modern Physics 74, 825 (2002).

[92] E. H. Wichmann and J. H. Crichton, Physical Review 132, 2788 (1963).

[93] J. Ellis, J. S. Hagelin, D. Nanopoulos, and M. Srednicki, Nuclear Physics B 241, 381 (1984).

[94] W. G. Unruh and R. M. Wald, Reports on Progress in Physics 80, 092002 (2017), arXiv:1703.02140 .

[95] W. Unruh and R. Wald, Physical Review D 52, 2176 (1995), arXiv:9503024v3 [arXiv:hep-th] .

[96] W. G. Unruh, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 370, 4454 (2012), arXiv:1205.6750 .

[97] J. Maldacena, Advances in Theoretical and Mathematical Physics 2, 231 (1998).

[98] L. Susskind, L. Thorlacius, and J. Uglum, Physical Review D 48, 3743 (1993).

[99] G. 't Hooft, Nuclear Physics B 335, 138 (1990).

[100] J. D. Bekenstein, Physical Review D 9, 3292 (1974), arXiv:arXiv:1011.1669v3 .

[101] B. Schumacher and M. Westmoreland, *Quantum Processes, Systems, and Information* (Cambridge University Press, Cambridge, 2010).

[102] D. A. Lowe, J. Polchinski, L. Susskind, L. Thorlacius, and J. Uglum, Physical Review D 52, 6997 (1995).

[103] N. Itzhaki, , 1 (1996), arXiv:9607028 [hep-th] .

[104] A. Giveon and N. Itzhaki, Journal of High Energy Physics 2012, 94 (2012).

[105] R. Bousso, Physical Review D 87, 124023 (2013).

[106] J. Preskill, , 16 (1992), arXiv:9209058 [hep-th] .

[107] S. W. Hawking, Communications in Mathematical Physics 87, 395 (1982).

[108] S. B. Giddings, Physical Review D 46, 1347 (1992), arXiv:9203059 [hep-th] .

[109] D. Garfinkle, G. T. Horowitz, and A. Strominger, Physical Review D 43, 3140 (1991).

[110] C. G. Callan, S. B. Giddings, J. A. Harvey, and A. Strominger, Physical Review D 45, R1005 (1992).

[111] A. Ashtekar, F. Pretorius, and F. M. Ramazanoğlu, 044040, 1 (2010), arXiv:1012.0077 .

[112] T. Banks, A. Dabholkar, M. R. Douglas, and M. O'Loughlin, Physical Review D 45, 3607 (1992), arXiv:9201061 [hep-th] .

[113] S. B. Giddings and W. M. Nelson, Physical Review D 46, 2486 (1992).

[114] D. N. Page, Physical Review D 13, 198 (1976).

[115] D. N. Page, Physical Review Letters 71, 1291 (1993), arXiv:9305007 [gr-qc] .

[116] S. K. Foong and S. Kanno, Physical Review Letters 72, 1148 (1994).

[117] J. Sánchez-Ruiz, Physical Review E 52, 5653 (1995).

[118] S. Sen, Physical Review Letters 77, 1 (1996).

[119] L. C. Malacarne, R. S. Mendes, and E. K. Lenzi, Physical Review E 65, 046131 (2002).

[120] E. Jaynes and F. Cummings, Proceedings of the IEEE 51, 89 (1963).

[121] A. D. Greentree, J. Koch, and J. Larson, Journal of Physics B: Atomic, Molecular and Optical Physics (2013), 10.1088/0953-4075/46/22/220201.

[122] K. Micadei, J. P. S. Peterson, A. M. Souza, R. S. Sarthour, I. S. Oliveira, G. T. Landi, T. B. Batalhão, R. M. Serra, and E. Lutz, Nature Communications , 3 (2017), arXiv:1711.03323 .

[123] P. A. Camati, J. P. Peterson, T. B. Batalhão, K. Micadei, A. M. Souza, R. S. Sarthour, I. S. Oliveira, and R. M. Serra, Physical Review Letters 117, 1 (2016).

[124] J. P. Peterson, R. S. Sarthour, A. M. Souza, I. S. Oliveira, J. Goold, K. Modi, D. O. Soares-Pinto, and L. C. Céleri, Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences 472 (2016), 10.1098/rspa.2015.0813.

[125] S. Weinberg and O. W. Greenberg, Physics Today (2008), 10.1063/1.2808256