IFT

Instituto de Física Teórica
Universidade Estadual Paulista

DISSERTAÇÃO DE MESTRADO

IFT-D.005/16

Information loss in black holes and the unitarity of quantum mechanics

Gabriel Cozzella

Orientador

G. E. A. Matsas

Julho de 2016

Acknowledgments

This work would not be possible without the endless sacrifice and love of my mother, Silvia, and the life-changing initial spark of interest for science given to me by my father, Marcos.

To my girlfriend, Raquel, for her support, patience and love. Without you my burdens would be much harder to carry. *Você é assim...*

To all my friends, who always had time to listen to my woes, a joke to tell, some advice to give, a problem to share and a way to make my life more worthwhile.

To my advisor, professor George E. A. Matsas, for the patience and attention dedicated to enlightening me throughout all our years together. Ευχαριστώ.

To FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo), for financial support under grant 2014/08684-9.

Abstract

The quantum theory of fields in curved space-times is the most solid framework for studying the interplay between gravity and quantum mechanics in the absence of a complete theory of quantum gravity. In this scenario, one problem that has drawn much attention from the theoretical physics community in the last decades is the so-called "black hole information loss paradox", where the evolution from an initial pure quantum state to a final mixed quantum state would constitute a violation of the laws of quantum mechanics. In this dissertation we argue that information loss does not violate quantum mechanics, being simply a consequence of the semi-classical framework adopted and that the question of information recovery needs to be addressed by a yet unknown theory of quantum gravity.

Key-words: Quantum field theory in curved space-times; Black holes; Information loss paradox.

Field of knowledge: Physics of elementary particles; Quantum field theory; General relativity.

Resumo

A teoria quântica de campos em espaços-tempos curvos é o arcabouço teórico mais sólido que temos para estudar a interação entre gravitação e mecânica quântica na ausência de uma teoria completa de gravitação quântica. Neste contexto, um problema que atraiu muita atenção dos físicos teóricos nas últimas décadas é o chamado "paradoxo da perda de informação em buracos negros", onde a evolução de um estado quântico puro inicial para um estado quântico misto final caracterizaria uma violação das leis da mecânica quântica. Nesta dissertação nós argumentamos que a perda de informação em si não viola as leis da mecânica quântica e é consequência direta da teoria semi-clássica utilizada. Finalmente, argumentamos que a questão da recuperação da informação deve ser tratada utilizando-se uma teoria de gravitação quântica ainda desconhecida.

Palavras-chaves: Teoria quântica de campos em espaços-tempos curvos; Buracos negros; Paradoxo da perda de informação.

Áreas do conhecimento: Física das partículas elementares; Teoria quântica de campos; Relatividade geral.

Contents

List of Figures

1 Introduction

It is well known that in the energy scales we can probe and observe today modern fundamental physics rests upon two major theoretical frameworks: General relativity and the quantum theory of fields. The successes of both theories in their respective domains of validity is astonishing. General relativity is verified to great precision in solar system tests (like light bending of distant stars by the Sun) and in extra-solar astrophysical settings (like in the period diminishing of binary pulsar systems), besides giving a model for the evolution of the universe which agrees with all available cosmological data up to this day [1]. Recent measurement of gravitational waveforms also represents one of the most stringent tests of general relativity to date, besides being the first measurement of general relativity in the strong non-linear domain [2]. Similarly, quantum field theory has been tested in particle accelerators routinely and provides us with some of the best agreements between theoretical prediction and experimental data in the history of mankind.

Despite these formalisms being around a century old, a satisfactory marriage of both has not been achieved yet and an elusive theory of quantum gravity remains the holy grail of modern theoretical physics. In this scenario one can argue that the most natural (and non-radical) way to study quantum effects in situations where gravity is not negligible is to use a semi-classical approach, which is called *quantum field theory in curved space-times* (QFTCS), provided that we are at energies far below the Planck scale. This framework arises in a similar spirit of the beginning days of modern quantum mechanics and studies of the interaction between the electromagnetic field and the hydrogen atom, where the atomic degrees of freedom were quantized, but those of the electromagnetic field were not. Being initially developed in the early 70s, QFTCS has not only provided a consistent way of studying black hole thermodynamics and inflation, but has also shone new light in the understanding of flat space-time quantum field theory (see, e.g., the Unruh effect).

The most famous effect predicted by QFTCS is undoubtedly Hawking radiation. Discovered in an epoch when researchers were trying to better understand the *laws of black hole mechanics* (nowadays recognized as simply the usual laws of thermodynamics applied to black holes), the study of radiation of evaporating black holes may be the best available theoretical laboratory today to understand something about quantum gravity.

In this setting, the so-called *black hole information loss paradox* plays an important role and has been a major research drive in this area during the last couple of decades.

One way to see how the "paradox" arises is to consider the evolution of a scalar quantum field $\hat{\phi}(x^a)$, which is initially in its vacuum state, in a space-time describing a collapsing star. The collapse will create a black hole, which in turn will evaporate due to the emission of thermal radiation. Assuming that the black hole evaporates completely, the final state of the field will be characterized by a mixed density matrix. This evolution seems to take a pure quantum state to a mixed quantum state, something that is not allowed in closed systems by quantum mechanical unitary evolution. Pointed out by many as a disastrous violation of quantum mechanics (which was responsible for the misnomer *paradox*, since the violation was discovered by applying the laws of quantum mechanics), it is recognized today that no violation of quantum theory arises in QFTCS. Framing this scenario and understanding why this is the case is the main point of this dissertation. We shall emphasize, however, that despite the semi-classical theory being entirely consistent, it is possible (but not necessary) that a theory of quantum gravity will "resolve" the black hole singularity and recover unitary evolution throughout the whole space-time. Understanding how this can be the case may (or may not) give clues about properties of quantum gravity.

In this work we proceed as follows. Chapter 2 is a review of the relevant concepts of general relativity that we will use, focusing mainly on the causal structure of space-times and initial-value problems for classical field theories. Chapter 3 is a brief review of relevant concepts of quantum mechanics, exposing concepts that may help us to better understand quantum evolution in general cases, namely density matrices and entanglement.

Special attention is dedicated to the *von Neumann entropy* as a way of measuring entanglement and to the concept of unitary evolution. Chapter 4 is concerned with the tools we will need from quantum field theory in curved space-times. We offer a self-contained introductory exposition of the canonical quantization of a scalar field in curved space-times and the renormalization of its energy-momentum tensor in special cases, which will enable us to better understand the process of black hole evaporation. A simple cosmological model and some aspects of an analog model of black hole evaporation utilizing accelerated mirrors are presented as applications of these concepts. Chapter 5 puts all previous tools to use in the theory of black hole thermodynamics by deriving the existence and spectrum of Hawking radiation, the final ingredient we need to finally pose the information loss scenario. Chapter 6 then presents an analysis of information loss in evaporating black hole space-times, arguing that there is no consistency problem in QFTCS. We also make a brief overview of some models that impose unitary evolution in the whole space-time and comment on their properties before ending with our conclusions and final remarks in chapter 7.

Throughout this work we follow sign conventions adopted in [3]. Therefore, $c = \hbar = k_B = G = 1$ unless explicitly noted. The metric signature is $(+, -, -, -)$, $R^a{}_{bcd} = \partial_d \Gamma^a{}_{bc} - ...$, $R_{ab} = R^c{}_{acb}$ and $G_{ab} = -8\pi T_{ab}$. In the sign convention of Misner, Thorn & Wheeler these definitions mean that we are $(-, -, -)$. Latin indexes $(a, b, ...)$ are used to denote tensorial quantities in arbitrary coordinate systems and Greek indexes $(\mu, \nu ...)$ to denote their value when a particular kind of coordinate system has been specified.

2 A brief review of general relativity

In this chapter we review and summarize some concepts about the causal structure of space-times. In particular, we focus on the Schwarzschild metric as an archetypal example of the causal structure one finds when dealing with black holes. We also review how the causal structure influences the well-posedness of initial value problems in classical field theory in curved space-times. Although we will be mainly interested in quantum field theory, this simpler setting will suffice to create intuition around the main point of this work in later chapters.

2.1 Causal structure of space-times and energy conditions

General relativity can be succinctly described as a theory in which space-time is composed of a smooth manifold M endowed with a Lorentzian metric g_{ab} which couples to the energy and matter content present in space-time through Einstein's field equations

$$G_{ab} = R_{ab} - \frac{1}{2}Rg_{ab} = -8\pi T_{ab}. \tag{2.1}$$

The metric is one of the two fundamental objects of the theory, describing how matter will behave in the space-time. The second fundamental feature is the topology of the manifold, which is not determined by the field equations alone. As is well known from special relativity, the causal relation between events in a space-time is determined by the behavior of its light-cones. Given two events $Q, P \in M$ we can only tell their causal relation if we know their corresponding light-cones. These, in turn, are determined by the null geodesics of the metric, i.e., curves $x^a(\lambda) \in M$ satisfying

$$g_{ab}\frac{\mathrm{d}x^a}{\mathrm{d}\lambda}\frac{\mathrm{d}x^b}{\mathrm{d}\lambda} = 0. \tag{2.2}$$

where $\lambda \in \mathbb{R}$ is an affine parameter. It is worthy to note that conformally related metrics give rise to the same causal structure.

A metric g'_{ab} is said to be *conformally related* to another metric g_{ab} if there exists a smooth positive non-vanishing function $\Omega(x^c)$ such that $g'_{ab} = \Omega^2(x^c) g_{ab}$. The fact that the causal structure is the same for conformally related metrics makes possible to depict space-time causal structures in a very useful way, namely the Carter-Penrose diagrams. This is achieved by a coordinate transformation that isolates a factor in the metric that diverges when we go to the asymptotic infinite regions. Multiplied then by an appropriate conformal factor, this renders the coordinate ranges finite. Since we will make extensive use of these diagrams in the rest of the work we give a quick summary of how they are constructed with the simplest example possible, Minkowski space-time, $(\mathbb{R}^4, \eta_{ab})$. Starting with spherical coordinates, where the metric has the form $diag(1, -1, -r^2, -r^2 \sin^2\theta)$, we can make the following coordinate transformations

$$u = t - r, \tag{2.3}$$

$$v = t + r, \tag{2.4}$$

$$U = 2 \arctan u, \tag{2.5}$$

$$V = 2 \arctan v, \tag{2.6}$$

where, since $u, v \in (-\infty, +\infty)$, $U, V \in [-\pi, \pi]$. Considering only the $t - r$ plane of the metric (i.e., keeping θ and ϕ fixed), the invariant interval takes the form

$$\mathrm{d}s^2 = \frac{1}{4} \sec^2(U/2) \sec^2(V/2)\ \mathrm{d}U \mathrm{d}V. \tag{2.7}$$

We can clearly identify the conformal factor $\Omega^{-2}(U, V) = \frac{1}{4} \sec^2(U/2) \sec^2(V/2)$ in the above expression. The conformally related length element is then simply $\mathrm{d}s' = \mathrm{d}U \mathrm{d}V$. Since the coordinates U and V now have a finite range we can draw the whole conformally related space-time as in figure 1, where each point represents a 2-sphere.

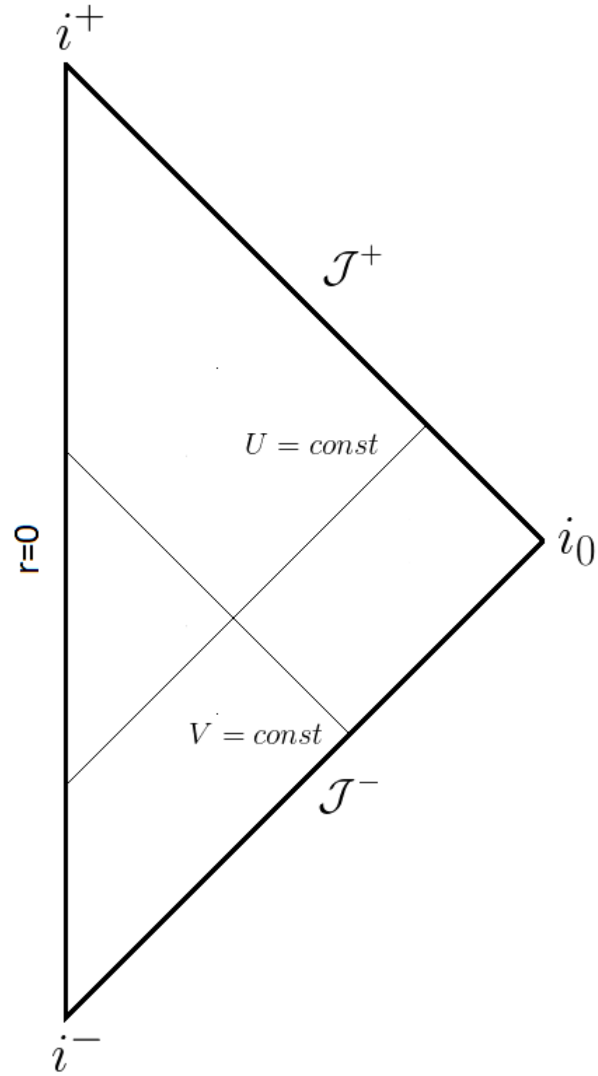

Figure 1: Carter-Penrose diagram of Minkowski space-time.

There are several regions of interest in this picture. Firstly, we have the points labelled i_0, i^+ and i^-. These are called *space-like infinity, future time-like infinity* and *past time-like infinity*, respectively. All space-like surfaces represented in this figure begin at $r = 0$ (the vertical left line) and end at i_0 ($r \to \infty$). All time-like geodesic trajectories begin at i^- and end at i^+. We also have $\mathcal{J}^+$ and $\mathcal{J}^-$, labelling the *light-like future* and *light-like past*, respectively. All light rays begin at $\mathcal{J}^-$ and end at $\mathcal{J}^+$ (reflecting at $r = 0$). The same is valid for asymptotically light-like behavior of time-like curves. The causal structure of Minkowski space-time is the simplest possible one in general relativity.

Since we will deal with more complex ones (mainly those of stationary and evaporating black holes) and since a good understanding of causal structures is paramount to the arguments laid out in this work we will review a bit more thoroughly concepts related to causal behavior. For more details, see, e.g., [4] or [5].

Given the wide variety of space-times that are solutions of Einstein's equations we start by assuming that we can make a continuous definition of future and past light-cones at each event P of a given space-time (M, g_{ab}). This guarantees that a smooth non-vanishing time-like vector field exists in such space-time. In Minkowski space-time, for example, this vector can be taken to be $t^a = (1, 0, 0, 0)$. Such a space-time is called *time-orientable*. Now, given a curve $\gamma(\lambda) \in M$, parameterized by λ, we say that

- $\gamma(\lambda)$ is a *time-like* future (past) directed curve if at each point $P \in \gamma(\lambda)$ the tangent vector to the curve at P, u^a, is time-like and is inside the future (past) light-cone of P.
- $\gamma(\lambda)$ is a *causal* future (past) directed curve if at each point $P \in \gamma(\lambda)$ the tangent vector to the curve at P, u^a, is either time-like or light-like and is in the future (past) light-cone of P.

With these definitions we can start describing causal relations between two events in M. Given two events $P, Q \in M$ we say that

- Q is in the chronological future of P if there exists an everywhere time-like future directed curve $\gamma(\lambda)$ such that $\gamma(0) = P$ and $\gamma(1) = Q$. The set of all such Qs is called the *chronological future* of P and is denoted by $I^+(P)$.
- Q is in the causal future of P if there exists a causal future directed curve $\gamma(\lambda)$ such that $\gamma(0) = P$ and $\gamma(1) = Q$. The set of all such Qs is called the *causal future* of P and is denoted by $J^+(P)$.

We define the chronological (causal) future of a set S by $I^+[S] = \bigcup_P I^+[P]$ ($J^+[S] = \bigcup_P J^+[P]$), where the union is taken over all points $P \in S$.

The concepts of chronological and causal past of events and sets of events are defined in a similar way and we denote them by $I^-(J^-)$. Physically, $I^+(S)$ consists of all points of space-time that can be influenced by signals that travel slower than light originating in S and $J^+(S)$ consists of points that can be influenced by signals that travel at a speed lower or equal than that of light. Similar ideas apply to $I^-(J^-)$.

When considering initial value problems in classical mechanics we need "an instant of time" in which initial conditions are given. These in turn, along with the equations of motion, suffice to determine the evolution of the dynamical system in consideration. To consider the evolution of fields in general curved space-times we need a similar idea of "an instant of time" that along with the equations of motion allow us to determine the field in the future. However, unlike the Newtonian case, more care is required since this should be done in a fully covariant and causal way. Intuitively we expect that the collections of events composing the initial instant of time should not have "time passing" between them. To construct a precise definition of this idea we define *achronal* sets as follows.

- We say that a set $S \subset M$ is *achronal* if for any given two points $P, Q \in S$ there is no time-like curve connecting P and Q, i.e., $Q \notin I^+(P)$. In other words, $I^+(S) \cap S = \emptyset$

An example of such set can be taken as the surfaces of constant t, where t is the usual Cartesian time coordinate, in Minkowski space-time. Now, to define how the initial conditions influence the system in the future we also need the idea of *domains of dependence*. This idea is not important in the Newtonian context since the speed of propagation of interactions is not an issue there. However, in relativistic theories we expect that a local change in the initial conditions should not change instantly the field since information takes a finite time to travel from point to point. To define the domains of dependence of a set S we also need the concept of inextendible curves. A curve $\gamma(\lambda)$ is said to be *inextendible to the future* if it contains no future end-point (i.e., $P \in M$ such that exists λ_0 such that for $\forall \lambda > \lambda_0$, $\gamma(\lambda)$ is contained in an open neighborhood of P). A similar definition can be made for past inextendibility. With this we are ready to define the different domains of dependence of a set S.

- The *future domain of dependence* of a set S, $D^+(S)$ is the set of points $P \in M$ such that every past directed inextendible causal curve through P crosses S at some point $Q \in S$.

- The *past domain of dependence* of a set S, $D^-(S)$ is the set of points $p \in M$ such that every future directed inextendible causal curve through P crosses S at some point $Q \in S$.

- The *domain of dependence* of a set S, $D(S)$ is then given by $D(S) = D^+(S) \cup D^-(S)$.

Physically, the domain of dependence of a set is the collection of events that are totally determined by data in S in a causal way. With all these ideas in place we can define *Cauchy surfaces*, which will correspond to the usual Newtonian idea of an instant of time and where we can define initial conditions to our initial value problems.

- A set Σ is called a *Cauchy surface* if it is a closed, achronal set for which $D(\Sigma) = M$.

It can be proven that Cauchy surfaces are embedded sub-manifolds in M that are three-dimensional, which justifies the intuitive interpretation of Σ as an instant of time [4]. Space-times that contain a Cauchy surface are called *globally hyperbolic*. These space-times have an array of nice causal properties and for this reason are the ones most considered when formulating quantum field theory in curved space-times. Among the most important of these properties is that such space-time manifold can be decomposed as $\mathbb{R} \times \Sigma_t$, where Σ_t is a family of Cauchy surfaces parameterized by t.

Since the causal structure of space-time is dependent on the energy-momentum content of such space-time, we need some knowledge of the behavior of the energy-momentum tensor that goes into the right-hand side of (2.1). To this regard, we remind the reader of the so-called *energy conditions*. Any symmetric and non-degenerate rank-2 tensor can be a solution of these equations for a suitably defined energy-momentum tensor, but most of these tensors are required to satisfy some conditions to exclude spurious non-physical solutions. The four conditions commonly required include:

- The *null energy condition (NEC)*, requiring that $T_{ab}k^a k^b \geq 0$ for any light-like vector k^a.

- The *weak energy condition (WEC)*, requiring that $T_{ab}t^a t^b \geq 0$ for any time-like vector t^a.

- The *strong energy condition (SEC)*, requiring $R_{ab}t^a t^b \geq 0$ for any time-like vector t^a.

- The *dominant energy condition (DEC)*, requiring $T^a_b t^b$ to be a time-like or light-like vector directed to the future for any time-like vector t^a directed to the future.

Such energy conditions appear in the proof of the area theorem and also in the proof of well-posedness of some initial-value problems in general relativity, which are relevant to this work.

Finally, we also remind the reader that a space-time is said to be *stationary* if it possess an asymptotically time-like Killing field[1] ξ^a. It is said to be *static* if besides being stationary the vector ξ^a satisfies $\xi_{[a}\nabla_b \xi_{c]} = 0$. An example of static space-time is Schwarzschild space-time and an example of stationary space-time is the Kerr space-time. Most space-times are neither static nor stationary, as, e.g., the space-time describing our universe.

2.2 The Schwarzschild metric

Undoubtedly the most "famous" (non-trivial) solution of Einstein's equations is the Schwarzschild solution. This space-time is the only spherically symmetric solution of Einstein's vacuum equation (a result known as Birkhoff's theorem) and its invariant length element can be expressed as

$$ds^2 = (1 - 2M/r)\, dt^2 - (1 - 2M/r)^{-1}\, dr^2 - r^2 d\theta^2 - r^2 \sin^2\theta\; d\phi^2, \tag{2.8}$$

[1] A vector field such that $\nabla_{(a}\xi_{b)} = \frac{1}{2}(\nabla_a \xi_b + \nabla_b \xi_a) = 0$.

where $M = const$ is the single parameter which characterizes it. The Schwarzschild metric, besides being spherically symmetric is also static. Despite appearing to have two "singularities" ($r = 2M$ and $r = 0$), only one of these is really a physical singularity, namely $r = 0$. This can be verified by computing scalar quantities like the Kretschmann scalar, $R^{abcd}R_{abcd} = \frac{48M^2}{r^6}$, which diverges for $r = 0$. Instead of being a singularity, $r = 2M$ characterizes a 3D null-surface called an *event horizon* (defined below). Due to the static nature of the Schwarzschild metric, the event horizon can also be characterized in a coordinate-free manner as the surface where $\xi^a \xi_a = 0$ with $\xi^a = \delta^a{}_0 \partial_t$.

To understand better the causal structure of the Schwarzschild metric we begin by finding its light-like coordinates in the $t-r$ plane, which will allow us to construct its Carter-Penrose diagram. Since for light-like geodesics $\mathrm{d}s^2 = 0$, we obtain, after integrating

$$t \pm \left[r + 2M \log\left(\frac{r}{2M} - 1 \right) \right] = const. \tag{2.9}$$

This motivates us to define *Eddington-Finkelstein* coordinates

$$v = t + r^*, \tag{2.10}$$

$$u = t - r^*, \tag{2.11}$$

where $r^* = r + 2M \log\left(\frac{r}{2M} - 1\right)$ is called the *tortoise radius*[2]. We make a final coordinate transformation with the intent of making explicit the conformal factor, namely

$$U = -e^{\frac{-u}{4M}}, \tag{2.12}$$

$$V = e^{\frac{v}{4M}}. \tag{2.13}$$

[2] Whose name derives from the fact that it takes an infinite coordinate range r^* to get from finite $r > 2M$ to $r = 2M$.

In terms of them the Schwarzschild length element (restricted to the $t-r$ plane) takes the form

$$\mathrm{d}s^2 = \frac{32M^3 e^{-r/2M}}{r}\mathrm{d}U\mathrm{d}V. \tag{2.14}$$

As we noted earlier, $r = 2M$ is not a true physical singularity of this space-time. This can be seen clearly from the length element written in the form above. Since there is no more coordinate singularity there we can extend the range of coordinates U and V from $(0, +\infty)$ to $(-\infty, +\infty)$ as long as $r > 0$. This allows us to eventually construct the *Kruskal maximal extension* of Schwarzschild space-time and draw its Carter-Penrose diagram.

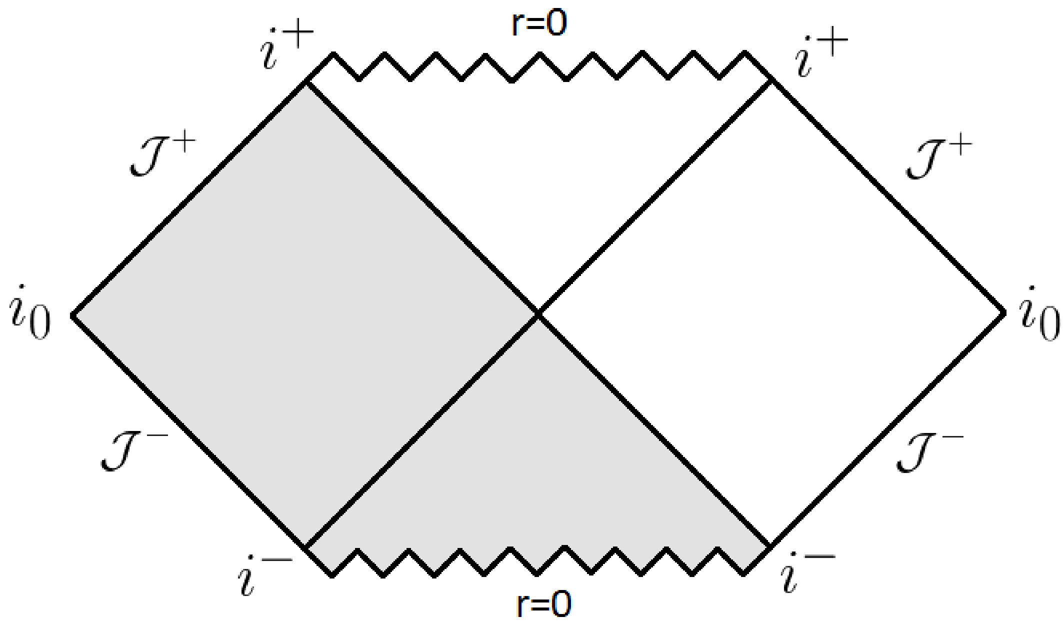

Figure 2: Carter-Penrose diagram of the maximal extension of Schwarzschild space-time. The coloring of regions are explained in the main text.

Although useful, only part of the full diagram (the white one) is physically relevant since the past singularity cannot be associated with the Big Bang. A more physically relevant situation that gives rise to a black hole is that of matter collapse (like what can happen in a massive star). Considering the beginning of the collapse at $t = -\infty$, the causal structure of this situation is given in figure 3.

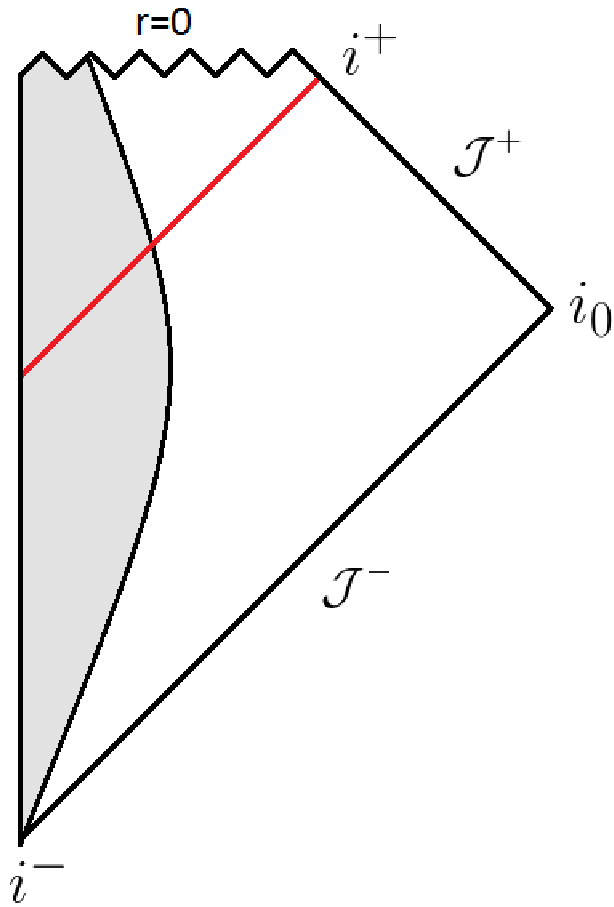

Figure 3: Carter-Penrose diagram of collapsing star forming a black hole. The gray region is the star interior, the red 45 degrees line in the middle is the event horizon and the wiggly line is the singularity.

The diagram 3 also motivates our definition of a black hole B in an asymptotically flat space-time. Since all light-like geodesics inside the region delimited by the red line hit the singularity, none of them escapes to $\mathcal{J}^+$. A *black hole* then is the set of points $B = M - J^-(\mathcal{J}^+)$ and its *event horizon*, H, is the boundary of B in M, $\partial B \cap M$.

An important property that we can define based on the time-like Killing vector ξ^a of the Schwarzschild space-time is its surface gravity, κ.

Physically speaking the surface gravity is the force someone at infinity ($r \to \infty$) must imprint at a unit mass particle near the horizon to keep it still (through an ideal rope, for instance). It is defined as

$$\kappa^2 = -\frac{1}{2}\nabla^a \xi^b \nabla_a \xi_b \Big|_H \tag{2.15}$$

The surface gravity, as we will see in later chapters of this work, will play the role of the black hole temperature (as measured by observers at infinity) when we take quantum effects into account. For Schwarzschild space-time, equation (2.15) reduces to $\kappa = 1/4M$.

There are many extensions of the simple solution we described above. A black hole can also possess angular momentum, electric charge or both. The Kerr, Reissner-Nordström and Kerr-Newman solutions characterize these cases, respectively. Black hole solutions can also be found for other theories of gravity with additional fields (like scalar-tensor theories) or different dimensionality. For the question of information loss, the Schwarzschild solution will suffice.

As can be seen comparing the Carter-Penrose diagrams 1 with 2 and 3, the causal structure of general space-times can be very different from the flat case. This in turn influences how one must proceed with defining well posed physical questions in arbitrary space-times. Great care regarding causality must be taken. We face this problem now.

2.3 Initial-value problems in curved space-times

Initial value problems in arbitrary space-times are more subtle than their Newtonian or special-relativistic counterparts. The reason, as we pointed out earlier, is that the causal structure of an arbitrary space-time can be quite different from that of Minkowski space-time. Given initial data on a Cauchy surface Σ_0 (which, as we saw, corresponds to our idea of "an instant of time") we want to evolve it using the equations of motion in such a way that the finite propagation speed of interactions is respected. Here focus is given to the problem of evolving a *classical* field in a globally hyperbolic region of space-time. We note that a good understanding of initial value problems in curved space-time is not only theoretically relevant, but is also essential for computational simulation of Einstein's equations.

We first begin by clarifying what is the meaning of a *well posed* initial value problem. We say that this is the case if both of the following criteria are met:

- Small changes in the initial data on Σ_0 should cause the new solution to be different from the old one by only a small amount (*continuity*).
- Changes in initial data on a subset of Σ_0, Σ_S, cannot modify the solution outside $J^+(\Sigma_S)$ (*causality*).

These definitions can be made mathematically precise, but this requires us to specify a proper theory. Since this is not yet the case, they are enough for the moment. Consider now a simple partial differential equation like the massless Klein-Gordon equation in Minkowski space-time, written in Cartesian coordinates

$$\frac{\partial^2 \phi}{\partial t^2} - \Delta \phi = 0, \tag{2.16}$$

where Δ is the Cartesian 3D Laplacian. If we are given both $\phi(t, \vec{x})$ and $\partial_t \phi(t, \vec{x})$ on a surface $t = t_0$ we can use the equation above to compute all time derivatives of order greater or equal than two. Assuming the data given is composed of analytical functions allows us then to write a power-series in t around t_0. It can be proven that this series has a finite radius of convergence, a result known as the *Cauchy-Kovalevskaya theorem*.

Unfortunately, this result is not enough to claim that our problem is well posed. We can see this by noting that the value of an analytical function at a point is determined by its values at an open neighborhood of this point. Therefore, changing the function at a point implies changing it in all points of Σ_0 instantly. Furthermore, the Cauchy-Kovalevskaya theorem has nothing to say regarding continuity (in the sense defined above) of the solution.

Nevertheless, not only it is possible to prove by other means the well-posedeness of equation (2.16), but to generalize such proof to a wider array of partial differential equations. We state the generalization, without proof[3], as in [4].

[3] A proof can be found in [5].

- Let (M, g_{ab}) be a globally hyperbolic region of an arbitrary space-time (possibly including the whole space-time), ∇_a a derivative operator and Σ a smooth, space-like Cauchy surface. Then, a system of partial differential equations of the form

$$g^{ab}\nabla_a\nabla_b\phi_i + \sum_j A^a_{ij}\nabla_a\phi_j + \sum_j B_{ij}\phi_j + C_i = 0, \tag{2.17}$$

  with $i, j \in \mathbb{N}$, has a well posed initial value formulation on Σ for smooth data $(\phi_i, n^a\nabla_a\phi_i)$, where n^a is normal to Σ.

The strongest hypothesis made is the statement above is that Σ is a Cauchy surface, mainly that its domain of dependence is the whole space-time region M. If this is not the case, there is no guaranteed way to specify uniquely the evolution of ϕ_i (although it may be possible in certain cases). Similarly, it is not possible to retro-evolve uniquely ϕ_i back to Σ from data on other space-like surfaces on M. Briefly speaking, *information was lost.* We will appeal to the intuition created by the last theorem when analyzing the behavior of quantum fields propagating in an evaporating black hole space-time. With these results we conclude our brief review of general relativity topics important to this work. Except for pointwise mentions, all we have done in this chapter omits quantum mechanics, to which we now turn.

3 A brief review of quantum mechanics

In this chapter we present some relevant concepts of quantum mechanics for analyzing processes involving quantum degrees of freedom, namely density matrices and entanglement. This is necessary since the final state of quantum fields in an evaporating black hole spacetime is a thermal state, which must be described by a mixed density matrix instead of a vector in some Hilbert space. Since this is ultimately due to the fact that degrees of freedom outside the horizon are entangled with those inside, we dedicate some attention to the von Neumann entropy as a tool to quantify entanglement. Finally, some simple results that are commonly used when discussing information loss are proved.

3.1 Density matrices

Let us consider an isolated system and its associated Hilbert space H_A. The first fundamental postulate of quantum mechanics tells us that an arbitrary *state* of the system is described by a linear, positive semi-definite[4] Hermitian operator with unity trace $\hat{\rho}$. This operator contains all available information about the system and generalizes the usual definition of a state as a ray in H_A. We say that the state $\hat{\rho}$ is *pure* if it can be written as $|\Psi\rangle\langle\Psi|$ for some $|\Psi\rangle \in H_A$, otherwise we say that the state is *mixed*. The expectation value of any observable $\hat{O}$ acting on H_A is defined as

$$\langle \hat{O} \rangle = Tr\left(\hat{\rho}\,\hat{O}\right), \tag{3.1}$$

where $Tr(\cdot)$ denotes the trace operation. Using the fact that $Tr\left(\hat{O}|\psi\rangle\langle\psi|\right) = \langle\psi|\hat{O}|\psi\rangle$ it is easy to see that when the state is pure this definition reduces to the usual one. If H_A has dimension $d_A = n$, we can pick an orthonormal basis $\{|i\rangle, i = 1, 2, ..., n\}$ and decompose $\hat{\rho}$ as

$$\hat{\rho} = \sum_{i,j}^{n} c_{ij} |i\rangle\langle j|, \tag{3.2}$$

[4] A positive semi-definite operator $\hat{A}$ is one such that $\langle\psi|\hat{A}|\psi\rangle \geq 0$ for every $|\psi\rangle \in H_A$.

with the condition $\sum_i^n c_{ii} = 1$ assuring its unit trace. Since the density matrix is Hermitian, the spectral theorem guarantees that it can be diagonalized. Denoting the orthonormal basis in which it is diagonal by $|d_i\rangle$, $\hat{\rho}$ can be written as

$$\hat{\rho} = \sum_i^n c'_{ii} |d_i\rangle\langle d_i|. \tag{3.3}$$

We note that the average value of the operator $\hat{P}_i = |d_i\rangle\langle d_i|$ in this state is just $Tr\left(\hat{P}_i\hat{\rho}\right) = c'_{ii}$. Therefore it is natural to interpret c'_{ii} as the probability of finding the system in the state $|d_i\rangle$ and denote it by p_i.

Another postulate of quantum mechanics assures us that there exists a unitary operator $\hat{U}(t, t_0)$ with $\hat{U}(t_0, t_0) = \hat{I}$, such that at any given time $t \geq t_0$ the density matrix of a *closed* system is given by

$$\hat{\rho}(t) = \hat{U}(t, t_0)\hat{\rho}(t_0)\hat{U}^\dagger(t, t_0). \tag{3.4}$$

The caveat *closed* is of fundamental importance. If the system is allowed to interact with another system, although the evolution of the composite system is still unitary (since this complete system is closed), evolution of any of its separate parts by themselves needs not to be. The same thing can happen if we only consider part of a physical system. Keeping this in mind will be of great help to understand why information is lost in semi-classical evaporating black holes.

If the system is part of a composite system a third postulate of quantum mechanics tells us that the Hilbert space of the full system is given by the tensor product of the Hilbert spaces of both systems. So, if system C is composed of systems A and B, we have $H_C = H_A \otimes H_B$. We can again choose orthonormal bases for both Hilbert spaces (with dimensions $d_A = n$ and $d_B = m$), $\{|i\rangle_A, |j\rangle_B\}$, and decompose any density matrix of system C as

$$\hat{\rho} = \sum_{ij}^n \sum_{kl}^m c_{ijkl} |i\rangle_A |k\rangle_B \langle j|_A \langle l|_B. \tag{3.5}$$

Composite systems constitute the majority of real world systems. They can be applied to describe open quantum systems, where the system of interest interacts with its environment, or to describe systems divided by some spatial boundary, be it arbitrary or imposed by nature (as is the case of black holes event horizons). In both cases we are only interested in what happens to part of the system and we need a consistent way of discarding information pertaining to the other part we are not interested (or do not have access). This is obtained by taking the *partial trace* of the full density matrix. A simple example of this can be seen by considering a bipartite system of two qubits in a Bell state[5], given by

$$|\Psi\rangle = \frac{1}{\sqrt{2}}\left(|01\rangle + |10\rangle\right), \tag{3.6}$$

whose density matrix is

$$\hat{\rho}_\Psi = |\Psi\rangle\langle\Psi| = \frac{1}{2}\left(|01\rangle\langle 01| + |01\rangle\langle 10| + |10\rangle\langle 01| + |10\rangle\langle 10|\right). \tag{3.7}$$

If someone makes a measurement of qubit B, then qubit A will be encountered in state $|0\rangle$ half of the times and in state $|1\rangle$ half of the times. The density matrix characterizing such situation is then

$$\hat{\rho}_A = \frac{1}{2}\left(|0\rangle\langle 0| + |1\rangle\langle 1|\right). \tag{3.8}$$

Defining the partial trace over system B as a linear operation such that

$$Tr_B\left(c_{ijkl}|ik\rangle\langle jl|\right) = c_{ijkl}\langle k|l\rangle|i\rangle\langle j|, \tag{3.9}$$

we can see that

$$Tr_B\left(\hat{\rho}_\Psi\right) = \hat{\rho}_A, \tag{3.10}$$

as expected.

[5]From now on, where there is no risk of confusion, we denote $|i\rangle_A|j\rangle_B$ as $|ij\rangle$.

Despite the state of the whole system being pure, the state just of system A is not. Mixed states can arise not only in this way (disregarding information about part of a composite system) but can also arise when we consider ensembles of different states with different probabilities. The most common example of this second type of mixed density matrix is a thermal state, which plays a fundamental role in thermodynamics and in the question of information loss. A thermal state has a density matrix given by

$$\hat{\rho}_\beta = \frac{1}{Z} e^{-\beta \hat{H}}, \tag{3.11}$$

where $Z = Tr\left(e^{-\beta \hat{H}}\right)$ is the partition function, $\hat{H}$ is the Hamiltonian of the system and $\beta = 1/T$, where T is the temperature. We can see that this is state is mixed by choosing a base $\{|\alpha\rangle\}$ in which $\hat{H}$ is diagonal, allowing us to write $\hat{\rho}$ as

$$\hat{\rho} = \frac{1}{Z} \sum_\alpha e^{-\beta E_\alpha} |\alpha\rangle\langle\alpha|, \tag{3.12}$$

with E_α being the eigenvalues of $\hat{H}$. Thermal states characterize black body radiation, among which Hawking radiation is the most relevant example for the purposes of this work.

3.2 The von Neumann entropy

Entanglement is one of the fundamental differences between classical and quantum states. Given a bipartite[6] system with Hilbert space $H_C = H_A \otimes H_B$ and a state $|\psi\rangle \in H_C$ we say that $|\psi\rangle$ is a *separable state* if it can be written as

$$|\psi\rangle = |\alpha\rangle \otimes |\beta\rangle, \tag{3.13}$$

where $|\alpha\rangle \in H_A$ and $|\beta\rangle \in H_B$. If this is not the case, then $|\psi\rangle$ is said to be an *entangled state*. An important question is how to quantify how much a state is entangled.

[6]Here we focus on bipartite states, since this case is the best understood one among all cases.

This is a complicated question for general mixed states [6]. We note that when dealing with black holes we will always choose an initial state which is pure and we will restrict ourselves to measuring entanglement of bipartite pure states. For this particular set of states, a useful way of quantifying entanglement is given by the von Neumann entropy, which is defined as

$$S(\hat{\rho}) = -Tr\left(\hat{\rho}\log\hat{\rho}\right). \tag{3.14}$$

The von Neumann entropy can be seen as the quantum version of Shannon entropy, a classical measure of information. An important property of the von Neumann entropy is that it vanishes if *and only if* $\hat{\rho}$ is a pure state. To see that $S(\hat{\rho})$ satisfies this property we start by writing $\hat{\rho}$ in a diagonal basis, obtaining

$$S(\hat{\rho}) = -\sum_{\alpha} p_{\alpha}\log p_{\alpha}, \tag{3.15}$$

where p_{α} are the eigenvalues of $\hat{\rho}$. When $\hat{\rho}$ is a pure state all p_{α} are 0 except for one which must be equal to 1 (to guarantee that $Tr(\hat{\rho}) = 1$).

Therefore $S(\hat{\rho}) = 0$[7]. On the other hand, if $S(\hat{\rho}) = 0$, since $-p_{\alpha}\log p_{\alpha} \geq 0$ for $p_{\alpha} \in [0, 1]$, we need to have all p_{α} equal to 0, except for one that needs to be equal to 1, due to the constraint that $\sum_{\alpha} p_{\alpha} = 1$. Therefore, $\hat{\rho} = |\alpha\rangle\langle\alpha|$ for some α. Another property of the von Neumann entropy concerns density matrices obtained by taking the partial trace of a pure state of a bipartite system. Consider again system C composed of system A and B and a pure state in H_C. By taking the partial trace with respect to A (B) we can obtain the reduced density matrix $\hat{\rho}_{A(B)}$. The von Neumann entropy is such that

$$S(\hat{\rho}_A) = S(\hat{\rho}_B). \tag{3.16}$$

This fact is a direct consequence of the Schmidt decomposition theorem for composite binary systems [7].

[7] Since $\lim_{x\to 0} x\log x = 0$.

Also, it is worth noting that this fact led to proposals of identifying entanglement entropy with black hole entropy. To see the motivation behind this idea, first consider a field $\hat{\phi}(x^a)$ in a pure state in Minkowski space-time $(\mathbb{R}^4, \eta_{ab})$. If we use spherical coordinates to subdivide space into two halves (e.g., $r < R$ and $r \geq R$ with some constant $R > 0$) and take the trace with respect to one of the halves we obtain a reduced density matrix with non-zero von Neumann entropy. However, we could make the same process by disregarding the other subsystem. Since both entropies need to be equal they must depend only on something shared by both systems, which, in this case, is the area of their boundary. Since entropy is adimensional we need some quantity with dimensions of area to make the dimensionality correct. A natural choice would be the Planck length squared, leading us to

$$S\left(\hat{\rho}_{A(B)}\right) \propto \frac{A}{l_p^2}. \tag{3.17}$$

A detailed calculation showing how this works can be found in [8] and [9]. In spite of being an artificial example (since the subdivision was done by hand), a real scenario in which this idea can be applied is in the Schwarzschild (or Kerr) space-time, in which the exterior region corresponds to the outside of the black hole event horizon and the interior region to the inside. A complete review about this application can be found in [10]. As we will see in chapter 5, this formula has the same functional form of the black hole thermodynamical entropy. Nevertheless, explaining black hole entropy as entanglement entropy has some problems dealing with gauge fields. Also, calculations involving entanglement entropy of quantum fields need to be regularized due to ultra-violet divergences. A natural scale for a cutoff is given by the Planck length, as seen above. Even if reasonable, this procedure can only be fully justified by a complete theory of quantum gravity.

Finally, we prove that the von Neumann entropy cannot change by unitary evolution. This will tell us unambiguously that a pure state (with $S(\hat{\rho}_{t=0}) = 0$) has evolved to a mixed state ($S(\hat{\rho}_{t>0}) \neq 0$). To do so, we prove that the characteristic polynomial P_λ of $\hat{\rho}_{t=0}$ and $\hat{\rho}_{t>0} = \hat{U}(t)\hat{\rho}_{t=0}\hat{U}^\dagger(t)$ are the same. Since the von Neumann entropy depends only on the eigenvalues of $\hat{\rho}$, as can be seen from equation (3.15), this is sufficient to show that it remains unchanged. We have

$$\begin{aligned} P_\lambda(\hat{\rho}_{t>0}) &= det(\hat{\rho}_t - I\lambda) && (3.18)\\ &= det(\hat{U}(t)\hat{\rho}_{t=0}\hat{U}^\dagger(t) - \hat{U}(t)\hat{U}^\dagger(t)\lambda) && (3.19)\\ &= det(\hat{U}(t)(\hat{\rho}_{t=0} - I\lambda)\hat{U}^\dagger(t)) && (3.20)\\ &= det(\hat{U}(t)\hat{U}^\dagger(t))det(\hat{\rho}_{t=0} - I\lambda) && (3.21)\\ &= det(\hat{\rho}_{t=0} - I\lambda) = P_\lambda(\hat{\rho}_{t=0}). && (3.22) \end{aligned}$$

Therefore, since $S(\hat{\rho}) = 0$ if and only if $\hat{\rho}$ is pure, this means that an initially pure state cannot evolve to a mixed state by unitary evolution. This is ultimately the fact responsible for sparking the whole initial discussion about the black hole information problem once the universe is considered a closed system.

With this, we finish our quick review of quantum theory and proceed to quantum field theory in curved space-times.

4 Quantum field theory in curved space-times

In this section we present a self-contained introductory exposition of (free) quantum field theory in curved space-times. The fundamental ingredients that go into quantizing a classical field theory in Minkowski space-time are discussed and the modifications needed to account for curved space-times are analyzed. We discuss a simple model of a scalar field in a homogeneous and isotropic space-time as an application of this quantization scheme, illustrating in a simple technical setting what kind of objects we will calculate in the next chapter.

The procedure of renormalization in curved space-time is then briefly presented (since renormalization is not trivial even for a free scalar field in curved space-time). This will allow us to better understand the flux of energy coming from an evaporating black hole space-time (and similar models). We end this chapter by showing some properties of a much studied model for black hole evaporation, which is the accelerated mirror model. It will be useful to keep these features in mind when discussing the real situation.

4.1 Canonical quantization in curved space-times

Consider a classical real scalar field $\phi(x^a)$ with mass m in Minkowski space-time. The Lagrangian density[8] of the field is given by

$$\mathcal{L} = \frac{1}{2}\left(\partial_a\phi\partial^a\phi - m^2\phi^2\right). \tag{4.1}$$

The field equation is obtained through the Euler-Lagrange equation, resulting in the (flat space-time) Klein-Gordon equation

$$\left(\partial_a\partial^a + m^2\right)\phi = 0. \tag{4.2}$$

[8]From this point on called simply Lagrangian.

Since $\partial_t, \partial_x, \partial_y$ and ∂_z are Killing fields in flat space-time, we try a solution of the form (utilizing the usual Cartesian coordinates)

$$\phi_{(\omega,\vec{k})}(x^\alpha) = \frac{e^{-i\omega t+\vec{k}\cdot\vec{x}}}{(2\pi)^3\sqrt{2\omega}} = \frac{e^{-ik_\mu x^\mu}}{(2\pi)^3\sqrt{2\omega}}. \tag{4.3}$$

By plugging this into the Klein-Gordon equation we obtain a solution if the usual dispersion relation $k^\mu k_\mu = m^2$ holds and $\omega = \sqrt{|\vec{k}|^2 + m^2} > 0$.

The solutions written in equation (4.3) are called *normal modes* of the field ϕ. They and their complex conjugates form a complete basis of solutions to the massive scalar wave equation, allowing us to express any solution as a linear combination of them. They also satisfy two important properties. Firstly, they are orthonormal with respect to the Klein-Gordon product

$$(f_1, f_2) = i\int \mathrm{d}\Sigma^a \left(f_2^* \overleftrightarrow{\partial_a} f_1\right) = i\int \mathrm{d}\Sigma^a \left(f_2^*\partial_a f_1 - f_1\partial_a f_2^*\right), \tag{4.4}$$

where Σ is a Cauchy surface. Secondly, the modes (4.3) are eigenfunctions of the operator $i\partial_t$ with *positive* eigenvalues. For this reason, they are called *positive frequency* modes (their complex conjugates being *negative frequency* modes). At this point this is just terminology, but it will be very useful below when we discuss quantization of the field theory in curved space-times, since these positive frequency modes are associated with the particle creation operators of the field. Since equation (4.2) is linear we can use (4.3) to construct its general solution. Imposing that the field is real we obtain

$$\phi(x^\alpha) = \int \frac{\mathrm{d}^3\vec{k}}{(2\pi)^3\sqrt{2\omega}} \left(a_{\vec{k}} e^{-ik_\mu x^\mu} + a_{\vec{k}}^* e^{ik_\mu x^\mu}\right). \tag{4.5}$$

From this we can obtain the canonical conjugate momentum

$$\Pi(x^\alpha) = \frac{\partial\mathcal{L}}{\partial(\partial_t\phi)} = \partial_t\phi = \frac{1}{i}\int \frac{\mathrm{d}^3\vec{k}}{(2\pi)^3}\sqrt{\frac{\omega}{2}} \left(a_{\vec{k}} e^{-ik_\mu x^\mu} - a_{\vec{k}}^* e^{ik_\mu x^\mu}\right). \tag{4.6}$$

The process of quantization consists of promoting the fields $\phi(x^\alpha)$ and $\Pi(x^\alpha)$ to operators $\hat{\phi}(x^\alpha)$ and $\hat{\Pi}(x^\alpha)$ by transforming $a_{\vec{k}}$ and their complex conjugates into operators acting on states of a Hilbert space H. We also impose the canonical commutation relations at equal times

$$[\hat{\phi}(t,\vec{x}),\hat{\phi}(t,\vec{y})] = [\hat{\Pi}(t,\vec{x}),\hat{\Pi}(t,\vec{y})] = 0, \tag{4.7}$$

$$[\hat{\phi}(t,\vec{x}),\hat{\Pi}(t,\vec{y})] = i\delta^3\left(\vec{x}-\vec{y}\right), \tag{4.8}$$

which imply the commutation relations for $\hat{a}_{\vec{k}}$

$$[\hat{a}_{\vec{k}},\hat{a}_{\vec{k}'}] = [\hat{a}^\dagger_{\vec{k}},\hat{a}^\dagger_{\vec{k}'}] = 0, \tag{4.9}$$

$$[\hat{a}_{\vec{k}},\hat{a}^\dagger_{\vec{k}'}] = \delta^3\left(\vec{k}-\vec{k}'\right). \tag{4.10}$$

These are nothing more than the commutation relations satisfied by an infinite set of quantum harmonic oscillators. It is then natural to interpret $\hat{a}_{\vec{k}}$ (and their conjugates) as annihilation (creation) operators. This implies, crucially, that they define a vacuum state, $|0\rangle$, that is annihilated by all $\hat{a}_{\vec{k}}$

$$\hat{a}_{\vec{k}}|0\rangle = 0 \quad \forall\vec{k}. \tag{4.11}$$

From this point on what it is usually done in Minkowski space-time quantum field theory is to show that the excitations of the vacuum state correspond to particles with a well-defined momentum and mass, build the Fock space, turn on interactions and calculate amplitudes using Feynman diagrams for this and more sophisticated theories (e.g., electromagnetism, Yang-Mills theories). This is the process that (modulo great technical complications) leads to some of the most experimentally successful predictions of modern physics alluded in chapter 1. The existence of a well-defined vacuum state is essential for the normal interpretation of excitations of the vacuum as particles. As we will see now, when gravity is turned on this ceases to be so simple. To this end we follow closely [11].

If we take a closer look into the process we used to quantize the scalar field theory above we can see that in order to quantize it we needed to do basically four things:

- Specify a Lagrangian;
- A procedure of quantization (here we used the canonical one, but we could just as well have used path-integral quantization);
- Specify the Hilbert space of the theory (usually done by constructing the Fock space associated with the state defined by equation (4.11));
- Specify the observables of the theory.

The two former concepts are applicable in all kinds of space-times, but the two latter ones are not. To pick a specific vacuum state we implicitly used Lorentz invariance (by adopting normal modes which are invariant by Lorentz transformations) as a preferred criterion. This invariance is *not* available in general space-times. This means that *there is no preferred criterion to choose between different unitarily inequivalent field representations*[9], which in turn means that different observers do not need to agree about the particle content of the theory. This is one of the main points we need to address in constructing a quantum field theory in curved space-times.

Firstly, we generalize the Lagrangian to make it compatible with general covariance. The only extra criterion we must fulfill is that the general Lagrangian reduces to the Minkowski one in the limit $g_{ab} \to \eta_{ab}$. Let us assume a Lagrangian of the form

$$\mathcal{L}_{g_{ab}} = \frac{\sqrt{-g}}{2}\left(\nabla_a\phi\nabla^a\phi - m^2\phi^2 - \xi R\phi^2\right), \tag{4.12}$$

where $\xi = const \in \mathbb{R}$ and R is the Ricci scalar. At this point we require that g_{ab} be asymptotically flat at i^-, i^+, $\mathcal{J}^-$ and $\mathcal{J}^+$.

[9] In fact, this can be already seen in flat space-time when we consider non-inertial observers (the Unruh effect).

This assumption (satisfied by all space-times used in this work) allows us to construct a set of complete normal modes in the time/light-like past, in the time/light-like future and expand the field at these regions in the same way we did for Minkowski space-time. The crucial point is that these two sets are not required to be equal, which in turn means that each set will define its own vacuum state. We will first lay down the general framework of expanding the field into two sets of different normal modes and apply it to a simple cosmological space-time as an example.

Consider two sets of positive-frequency normal modes in the time- or light-like past and time- or light-like future, f_i and F_i, where i labels all the relevant quantum numbers for the field. We note two important things: firstly, the labelling indexes i do not need to be discrete. Secondly, the modes are solutions of the equations of motion *at all points in space-time*, but assume the form of flat space-time normal modes at the asymptotic regions. Requiring them to be orthonormal with respect to the Klein-Gordon product (4.4) means that they satisfy

$$(f_i, f_j) = (F_i, F_j) = \delta_{ij}, \tag{4.13}$$

$$(f_i^*, f_i^*) = (F_i^*, F_i^*) = -\delta_{ij}, \tag{4.14}$$

$$(f_i, f_j^*) = (F_i, F_j^*) = 0. \tag{4.15}$$

We can expand the field using both sets of modes as

$$\phi = \sum_i \left(a_i f_i + {a_i}^* {f_i}^* \right), \tag{4.16}$$

$$\phi = \sum_i \left(b_i F_i + {b_i}^* {F_i}^* \right). \tag{4.17}$$

Since both sets are complete we can expand the modes of one in terms of the modes of the other, giving us

$$F_i = \sum_j \left(\alpha_{ij} f_j + \beta_{ij} f_j^*\right), \tag{4.18}$$

$$f_i = \sum_j \left(\gamma_{ij} F_j + \delta_{ij} F_j^*\right). \tag{4.19}$$

Imposing consistency with (4.13), (4.14) and (4.15) we see that the expansion coefficients $\alpha_{ij}, \beta_{ij}, \gamma_{ij}$ and δ_{ij} are not all independent and satisfy the relations

$$\sum_j \left(\alpha_{ij}\alpha_{kj}^* - \beta_{ij}\beta_{kj}^*\right) = \delta_{ik}, \tag{4.20}$$

$$\sum_j \left(\alpha_{ij}\beta_{kj} - \beta_{ij}\alpha_{kj}\right) = 0, \tag{4.21}$$

$$\gamma_{ij} = \alpha_{ji}^*, \tag{4.22}$$

$$\delta_{ij} = -\beta_{ji}. \tag{4.23}$$

The independent coefficients α_{ij} and β_{ij} are called *Bogolubov coefficients*. They describe how a positive-frequency mode in the past f_i is seen by a future observer as a mixture of positive and negative-frequency modes in the future. Now, quantizing the field, we promote the a_i and b_i (and their complex conjugates) to annihilation (creation) operators in the past and future, respectively. By equating (4.16) with (4.17) and using (4.18) and (4.19), we can obtain a relation between the creation and annihilation operators of the past and future, given by

$$\hat{b}_i = \sum_j \left(\alpha_{ij}^* \hat{a}_j - \beta_{ij}^* \hat{a}_j^\dagger\right), \tag{4.24}$$

$$\hat{b}_i^\dagger = \sum_j \left(-\beta_{ij} \hat{a}_j + \alpha_{ij} \hat{a}_j^\dagger\right). \tag{4.25}$$

As described earlier, a set of annihilation and creation operators define a vacuum state. But, it can be seen from the equations above that the vacuum state defined by the $\hat{a}_i$ operators cannot be a vacuum state of the $\hat{b}_i$ operators *if* $\beta_{ij} \neq 0$.

A simply consequence of this is that observers in the past and in the future will disagree on the presence or absence of particles. Consider the field initially in the past vacuum state, which from now on we will label as $|0\rangle_{in}$[10]. The number operator of mode i in the future, $\hat{N}_i = \hat{b}_i^\dagger \hat{b}_i$, has the expectation value

$$\left\langle \hat{N}_i \right\rangle = {}_{in}\langle 0|\hat{b}_i^\dagger \; \hat{b}_i|0\rangle_{in} = \sum_j |\beta_{ij}|^2. \tag{4.26}$$

This last relation perhaps illustrates one of the most important lessons that can be learned from QFTCS, the fact that *the particle content of a theory is observer dependent.*

We will make use of the formalism outlined above throughout the rest of this work, using it to describe the behavior of fields in the presence of changing boundary conditions (moving mirrors) and of a non-stationary collapsing star space-time (which will give rise to the Hawking effect).

Firstly, to illustrate this formalism, we will consider first a field propagating in a spatially flat homogeneous and isotropic space-time with an asymptotic constant scale factor $a(\eta)$ both in the past and in the future[11]. We can write the metric of this space-time as

$$\mathrm{d}s^2 = a^2(\eta)\left(\mathrm{d}\eta^2 - \mathrm{d}x^2 - \mathrm{d}y^2 - \mathrm{d}z^2\right). \tag{4.27}$$

Assuming the Lagrangian of equation (4.12) gives us the wave-equation

$$(\nabla_\mu \nabla^\mu + m^2 + \xi R)\phi = \frac{1}{a^4(\eta)}\partial_\eta[a^2(\eta)\partial_\eta \phi] - \frac{1}{a^2(\eta)}(\Delta_\perp \phi) + (m^2 + \xi R)\phi = 0, \tag{4.28}$$

where $\Delta_\perp$ is the Cartesian 3D Laplacian. Since the metric is both homogeneous and isotropic

[10] Similarly, the future vacuum state will be labelled $|0\rangle_{out}$.

[11] Given our universe, a more natural requirement would be an asymptotic constant (or slowly-varying) scale factor only in the future. This can be done at the cost of adding more technical burden, which escapes the scope of this example [12]. Here we just consider this space-time as a toy model.

we can try an *ansatz* to the above equation of the form

$$f_{\vec{k}} = \frac{g(\eta)}{a(\eta)\sqrt{2\pi^3}} e^{i\vec{k}\cdot\vec{x}}. \tag{4.29}$$

This reduces the wave-equation to an ordinary differential equation of a harmonic oscillator with a η-dependent spring frequency

$$\frac{\mathrm{d}^2 g(\eta)}{\mathrm{d}\eta^2} + [k^2 - V(\eta)]g(\eta) = 0, \tag{4.30}$$

where $V(\eta) = -a(\eta)^2 \left[m^2 + (\xi - 1/6)R(\eta)\right]$[12]. To guarantee normalization according to (4.13) $g(\eta)$ must also satisfy

$$g(\eta)^* \partial_\eta g(\eta) - g(\eta)\partial_\eta g(\eta)^* = -i. \tag{4.31}$$

Since in the infinite past $V(\eta) \to -a(-\infty)^2 m^2 = const$ we will pick $f_{\vec{k}}$ to be positive-frequency modes in the past region. Boundary conditions for $g(\eta)$ must be specified such that $f_{\vec{k}}$ behave like $e^{-ik_\mu x^\mu}$ in this region. Therefore we impose that

$$g(\eta) \to \frac{1}{\sqrt{2\omega}} e^{-i\omega\eta} \quad as \quad \eta \to -\infty, \tag{4.32}$$

where $\omega = \omega(\vec{k}, m) > 0$ is the relevant dispersion relation. Now, a precise solution of (4.28) is not known for an arbitrary scale factor. Some highly interesting simple examples that have analytical solutions can be seen, for instance, in [12] and [13]. Here we will not specify a precise form for $a(\eta)$ but just assume that it gives a $V(\eta)$ such that $|V(\eta)| \ll 1$. This approximation allow us to solve (4.30) iteratively by using the Green function method.

Firstly, we recast (4.30) as an integral equation already embodying the boundary condition (4.32), writing it as

$$g(\eta) = \frac{1}{\sqrt{2\omega}} e^{-i\omega\eta} + \frac{1}{\omega} \int_{-\infty}^{\eta} \mathrm{d}\eta' \; V(\eta') \sin\left(\omega(\eta - \eta')\right) g(\eta'). \tag{4.33}$$

[12]Here we have used that $\frac{\partial_\eta^2 a}{a^3} = R/6$.

Using our hypothesis that $|V(\eta)| \ll 1$ we substitute $g(\eta)$ inside the integral by the form that $g(\eta)$ has in past infinity, allowing us write a closed form integral for Bogolubov coefficients. This can also be viewed as a (first order in $V(\eta)$) perturbation expansion.

Applying this procedure to the integral equation above leads us to

$$g(\eta) = \frac{1}{\sqrt{2\omega}} e^{-i\omega\eta} + \frac{1}{\omega\sqrt{2\omega}} \int_{-\infty}^{\eta} \mathrm{d}\eta' \; V(\eta') \sin\left(\omega(\eta - \eta')\right) e^{-i\omega\eta'} \tag{4.34}$$

$$= \frac{1}{\sqrt{2\omega}} e^{-i\omega\eta} - \frac{i}{2\omega\sqrt{2\omega}} \int_{-\infty}^{\eta} \mathrm{d}\eta' \; V(\eta') \left(e^{i\omega(\eta-\eta')} - e^{-i\omega(\eta-\eta')} \right) e^{-i\omega\eta'} \tag{4.35}$$

$$= \frac{1}{\sqrt{2\omega}} e^{-i\omega\eta} \left(1 + \frac{i}{2\omega} \int_{-\infty}^{\eta} \mathrm{d}\eta' \; V(\eta') \right) - \frac{1}{\sqrt{2\omega}} e^{i\omega\eta} \left(\frac{i}{2\omega} \int_{-\infty}^{\eta} \mathrm{d}\eta' \; V(\eta') e^{-2i\omega\eta'} \right). \tag{4.36}$$

Taking $\eta \to \infty$, we can read directly the Bogolubov coefficients[13]

$$\alpha_\omega = \left(1 + \frac{i}{2\omega} \int_{-\infty}^{+\infty} \mathrm{d}\eta' \; V(\eta') \right), \tag{4.37}$$

$$\beta_\omega = -\left(\frac{i}{2\omega} \int_{-\infty}^{+\infty} \mathrm{d}\eta' \; V(\eta') e^{-2i\omega\eta'} \right). \tag{4.38}$$

Since β_ω will not be zero for arbitrary $V(\eta)$, this is an explicitly demonstration that in this space-time the in-vacuum and out-vacuum are not equal. An observer in the infinite future will then calculate a non-zero mean density of particles if the initial state of the field is $|0\rangle_{in}$.

It is noteworthy to say that we could only speak of particles in the infinite past and infinite future because we assumed asymptotically flatness. The particle content of space-time is *ill-defined* during the evolution[14]. The main protagonists of the quantum theory of fields are really *fields* and not particles. For this reason, it is useful to look at other observables different from the mean number of particles. It is to this task that we now turn.

[13] Modes with different ω's do not mix due to the symmetries of the problem.

[14] Naturally a reasonable definition of particles can be given for space-times that "do not change too fast" [12]. This is what allows us to use flat space-time quantum field theory every day at Earth laboratories.

4.2 Energy-momentum tensor renormalization

As is well known, the energy-momentum tensor is a fundamental component of any theory involving curved space-times for various reasons. Two important ones are that it is connected to the geometry of space-time by Einstein field equations and it is a local observable, in contrast to the mean particle number. In Minkowski space-time it can be defined via Noether's theorem using Lorentz invariance. We use a more general definition here well suited for curved space-times that has the additional advantage of giving a symmetric tensor without need of any symmetrization procedure. Consider the Einstein-Hilbert action together with the Lagrangian for matter fields

$$S = \frac{1}{16\pi} \int \mathrm{d}^4 x \sqrt{-g}\,(R + 16\pi\mathcal{L}) = S_{grav} + S_m. \tag{4.39}$$

By varying the action with respect to the metric and imposing it to be zero we obtain

$$0 = \frac{\delta S}{\delta g^{ab}} = -\left(R_{ab}(x) - \frac{1}{2}R(x)g_{ab}(x)\right) - \frac{16\pi}{\sqrt{-g(x)}}\frac{\delta(S_m)}{\delta g^{ab}(x)}, \tag{4.40}$$

where, throughout this section, x denotes a general space-time point. To make this consistent with Einstein field equations we *define* the energy-momentum tensor as the functional derivative of the matter action with respect to the metric

$$T_{ab}(x) = \frac{2}{\sqrt{-g(x)}}\frac{\delta(S_m)}{\delta g^{ab}(x)} = \int \mathrm{d}^4 x' \left(2\frac{\delta(\mathcal{L}(x'))}{\delta g^{ab}(x)} - g_{ab}(x')\mathcal{L}(x')\delta(x - x')\right). \tag{4.41}$$

If we apply it to the Lagrangian given in (4.12) (with $m = \xi = 0$, for simplicity) we obtain

$$T_{ab} = \nabla_a\phi\nabla_b\phi - \frac{1}{2}g_{ab}\nabla^c\phi\nabla_c\phi. \tag{4.42}$$

As we can see the energy-momentum tensor is composed of quadratic terms in relation to the fields. This is also the case if m or ξ are not zero.

When we quantize the field and try to calculate expectation values, these terms turn out to be divergent. For this reason, a renormalization procedure is required. In flat space-time, this process is usually carried out by the normal ordering prescription, where we define the operation of normal order, (: :), as

$$: \hat{a}\hat{a}^{\dagger} : = \hat{a}^{\dagger}\hat{a}, \tag{4.43}$$

$$: \hat{a}^{\dagger}\hat{a} : = \hat{a}^{\dagger}\hat{a}. \tag{4.44}$$

This eliminates the infinite zero point energy in the expectation value of the Hamiltonian for the quantized field and amounts to nothing more than simply throwing away a constant with infinite value. However, serious doubts may be raised on this prescription when we deal with curved space-times. This is due to the fact that what curves space-time is the absolute value of energy. Therefore, we are not free to re-scale the vacuum energy arbitrarily as in flat space-time. An idea to solve this would be to calculate the expected value of the energy-momentum tensor with an ultraviolet regulator, see which parameters need to be adjusted to recover Minkowski space-time, subtract the remaining terms from the original calculation and then take the limit where the regulator goes to infinity. This prescription also fails. An explicit example of such failure is given in [3]. Since these simple methods do not work we need more care when dealing with the energy-momentum tensor in curved space-times.

Although the renormalization machinery can be quite complicated, our main focus will be on the 2D case where things are a bit simpler. In fact, restricting ourselves to 2D gives us a major technical advantage since every 2D space-time is conformally flat. This will not hinder us in any way, since when analyzing the Hawking flux we will only care about the $t - r$ plane (effectively rendering the problem 2D). This is the most technical part of this work. No prejudice to its understanding will occur to the reader that only wishes to apply the results presented at the end of this section (equations (4.71) and (4.72)) without deriving them. Finally, we note that our treatment is mainly based on [3].

Our strategy is two-fold[15]: Firstly, we look for an *effective* action, W_{eff}, which would give us the semi-classical Einstein equations (where the expectation value of $\hat{T}_{ab}$ enters in the right-hand side of (2.1)) instead of the classical ones. Secondly, we will study the so-called *conformal anomaly*, which in the conformally flat case will determine completely the renormalized energy-momentum tensor.

Comparison with (4.41) lead us to look for W_{eff} such that

$$\frac{{}_{out}\langle 0|\hat{T}_{ab}|0\rangle_{in}}{{}_{out}\langle 0|0\rangle_{in}} = \frac{2}{\sqrt{-g}}\frac{\delta W_{eff}}{\delta g^{ab}}. \tag{4.45}$$

The reason for the appearance of the combination ${}_{out}\langle 0|\cdot|0\rangle_{in}$ instead of ${}_{in}\langle 0|\cdot|0\rangle_{in}$ or ${}_{out}\langle 0|\cdot|0\rangle_{out}$ (as can be expected from the semi-classical Einstein equation) is that our derivation is mainly based on the partition function, defined as

$$Z[0] = {}_{out}\langle 0|0\rangle_{in}. \tag{4.46}$$

In Minkowski space-time $Z[0]$ usually is normalized to one. This can no longer be the case in curved space-times, since the in and out vacuums differ in general. We note that our choice of initial and final states in the definition of the partition function $Z[0]$ will not change the results presented below. This is due to the fact that the divergences we treat here arise from the short-scale behavior of $\hat{T}_{ab}$ and this is independent of which combination of states $|0\rangle_{in}$ and $|0\rangle_{out}$ we pick.

To find the form of the effective action we will use the path-integral quantization of the scalar field. The usual partition function of the field is given by

$$Z[J] = \int D\phi \; e^{iS_m + i\int \mathrm{d}^4x'\phi(x')J(x')}. \tag{4.47}$$

where $J(x)$ is a scalar source current.

[15]At this part only we take $G \neq 1$. Therefore, objects can have dimensions of $(\text{mass})^n$. The reason for this will be addressed in the main text.

Since we are only dealing with free-field theories (coupled to gravity) we will set the source to zero. Consider now a variation of $Z[0]$ with respect to g^{ab}:

$$\frac{\delta Z[0]}{\delta g^{ab}} = i \int D\phi \, \frac{\delta S_m}{\delta g^{ab}} \, e^{iS_m}. \tag{4.48}$$

Multiplying both sides by $\frac{2}{\sqrt{-g}}$ we obtain

$$\frac{2}{\sqrt{-g}} \frac{\delta Z[0]}{\delta g^{ab}} = i \int D\phi \, \frac{2}{\sqrt{-g}} \frac{\delta S_m}{\delta g^{ab}} \, e^{iS_m} = i \; {}_{out}\langle 0|\hat{T}_{ab}|0\rangle_{in}. \tag{4.49}$$

Comparing the expression above with (4.45) motivate us to define the effective action by

$$Z[0] = e^{iW_{eff}}, \tag{4.50}$$

or, more explicitly

$$W_{eff} = -i \log\left(Z[0]\right) = -i \log\left({}_{out}\langle 0|0\rangle_{in}\right). \tag{4.51}$$

The partition function can be calculated from (4.47) giving us

$$Z[0] \propto \sqrt{det(-\hat{G}_F)}, \tag{4.52}$$

where $\hat{G}_F$ is the Feynmann Green function defined by

$$iG_F(x, x') = i\langle x|\hat{G}_F|x'\rangle = {}_{out}\langle 0|T\left[\hat{\phi}(x)\hat{\phi}(x')\right]|0\rangle_{in}, \tag{4.53}$$

T being the time-ordering operation. Substituting this form of $Z[0]$ in (4.51) we end up with

$$W_{eff} = -\frac{i}{2} \, Tr\left(\log\left(-\hat{G}_F\right)\right), \tag{4.54}$$

where $Tr(\cdot)$ stands for the trace operation[16].

[16] The proportionality constant in (4.52) can be absorbed in the definition of W_{eff}.

To give meaning to the formal expressions presented we will use the *generalized ζ-function* technique. Expanding $-\hat{G}_F$ in its eigenbasis we have

$$-\hat{G}_F = \sum_n \lambda_n^{-1} |n\rangle\langle n|, \tag{4.55}$$

where λ_n and $|n\rangle$ satisfy

$$-\hat{G}_F^{-1}|n\rangle = \lambda_n |n\rangle. \tag{4.56}$$

Consider now $(-\hat{G}_F)^\nu$. Since the eigenbasis is orthonormal (due to the properties of $\hat{G}_F^{-1}$), we can calculate easily its trace, giving our *definition* of the generalized ζ-function

$$\zeta(\nu) \equiv Tr(-\hat{G}_F)^\nu = \sum_n \lambda_n^{-\nu}. \tag{4.57}$$

Now, the argument of the logarithm in (4.54) needs to be dimensionless. To do so we insert a parameter μ with dimensions of (mass). In the end we will see that physical results do not depend on μ. The generalized ζ-function gives us

$$\begin{aligned} W_{eff} &= -\frac{i}{2} \lim_{\nu \to 0} Tr\ \partial_\nu \left(-\mu^2 \hat{G}_F\right)^\nu \\ &= \lim_{\nu \to 0} -\frac{i}{2} \mu^{2\nu} \left(\partial_\nu \zeta(\nu) + \zeta(\nu) \log \mu^2\right). \end{aligned} \tag{4.58}$$

The expression above does not necessarily converge for any value of ν. Nevertheless, it can be *defined* at these values by analytic continuation based on regions in which it converges. This analytic continuation is the process ultimately responsible for regulating the divergences present in the effective action[17].

[17] Some doubts arise about this process, but alternative (and more rigorous) methods can be used to show that it gives the correct result [3].

To evaluate (4.58) we need a representation of $\hat{G}_F$. Using the Schwinger-De Witt representation given in [3] allows us to write

$$\zeta(\nu) = i\Gamma(\nu)^{-1}(4\pi)^{-2}\int \mathrm{d}^4x\sqrt{-g}\int_0^{+\infty} \mathrm{d}(is)(is)^{\nu-3}e^{-im^2s}F(x,x,;is), \tag{4.59}$$

where $F(x,x';is) = \sum_{n=0}^{+\infty} a_n(x,x')(is)^n$ with $a_0(x,x') = 1$ and $a_n(x,x')$ being determined recursively from $a_0(x,x')$[18]. Using (4.59) we calculate (4.58), giving

$$\begin{aligned} W = &\int \mathrm{d}^4x\sqrt{-g}\left[-\frac{1}{64\pi^2}\int_0^{+\infty} \mathrm{d}(is)\log{(is)}\frac{\partial^3}{\partial(is)^3}\left(F(x,x;is)e^{-ism^2}\right)\right] \\ &+\int \mathrm{d}^4x\sqrt{-g}\left[\left(\frac{\log\mu^2}{32\pi^2}+\left(\frac{1}{64\pi^2}(\gamma-3/2)\right)\right)\left(\frac{1}{2}m^4-m^2a_1(x,x)+a_2(x,x)\right)\right], \end{aligned} \tag{4.60}$$

where γ is the Euler-Mascheroni constant. The first term is the renormalized action. The second contains only constants and geometrical terms[19]. These are the divergent terms which can be grouped together with the gravitational part of the action.

The bare infinite constants (like G and the cosmological constant Λ) in the gravitational action are then renormalized to their experimentally measured values[20]. Let us turn our attention now to conformal transformations[21]. By a conformal transformation[22] $g_{ab}(x) \to \Omega^2(x)g_{ab}(x)$ the classical action of the field changes by an amount $\delta S = S(g'_{ab}) - S(g_{ab})$ given by

$$\delta S = -\int \mathrm{d}^4x\sqrt{-g'}\; T_a{}^a\; \Omega^{-1}(x)\delta\Omega(x), \tag{4.61}$$

where $T_a{}^a = g_{ab}T^{ab}$ is the trace of the energy-momentum tensor defined previously.

[18] The explicit determination of the $a_n(x,x')$ does not concern us here. For details, the reader is referred to [3].

[19] $a_1(x,x)$ and $a_2(x,x)$ are composed only of geometrical quantities like the Riemann tensor and its contractions.

[20] Higher order bare terms may also be needed, but this does not concern us here.

[21] Due to our previous remarks that the divergences treated by renormalization are due to short-scale ultraviolet behavior (which is state-independent), $\langle\hat{T}_{ab}\rangle$ denotes the expectation value of $\hat{T}_{ab}$ in any combination of vacuum states. We emphasize that the numerical values may depend on the states, but the *expressions* involving $\langle\hat{T}_{ab}\rangle$ remain valid.

[22] This transformation is also called *Weyl transformation* to distinguish it from coordinate transformations that give a Ω^2 factor in front of the metric. Here we are not considering coordinate transformations.

If the action is invariant by the transformation $g_{ab}(x) \to \Omega^2(x) g_{ab}(x)$ the trace of the energy-momentum tensor vanishes. This can only be the case if there are no length-scales in the theory (like mass scales).

For this reason, we shall work in the massless limit. Now, conformal invariance of the *classical* action do not mean that the *quantum* energy-momentum tensor operator will have a zero trace. This is known as the *conformal anomaly*. Let us demonstrate this fact. Using the transformation properties of the Feynman propagator under conformal transformations the effective action transforms to

$$W' = -\frac{i}{2} Tr\left(\log\left(-(\mu\Omega)^2 \hat{G}_F\right)\right). \tag{4.62}$$

We also have

$$\langle \hat{T}_a{}^a(x) \rangle = -\frac{\Omega(x)}{\sqrt{-g}} \frac{\delta W'}{\delta \Omega(x)}, \tag{4.63}$$

where the last expression is to be calculated at $\Omega(x) = 1$. Note that $\Omega(x)$ only appears in the combination $\mu\Omega$ when we write W'. When the expression for the effective action (eq. (4.60) with $\mu \to \mu\Omega$) is applied to (4.63), we see that the finite term makes no contribution and we end up with the four-dimension conformal anomaly

$$\langle \hat{T}_a{}^a(x) \rangle_{4D} = -\frac{a_2(x,x)}{16\pi^2}, \tag{4.64}$$

proving that even if the initial classical action is conformally invariant the renormalized energy-momentum tensor has a non-zero trace. A similar calculation can be done in 2D, which gives the two-dimensional conformal anomaly

$$\langle \hat{T}_a{}^a(x) \rangle_{2D} = -\frac{a_1(x,x)}{4\pi} = -\frac{R}{24\pi}, \tag{4.65}$$

where we have used the explicit form of $a_1(x,x)$ in two dimensions. In the particular case of a conformally trivial theory (where space-time is conformally flat and the matter action is also conformal) this result will allow us to completely determine all components of $\langle \hat{T}_{ab} \rangle$. Consider a variation of W under a conformal transformation

$$W' = W - \int \mathrm{d}^4 x \sqrt{-g'} \langle \hat{T}_a{}^a \rangle \Omega^{-1}(x) \delta\Omega(x). \tag{4.66}$$

By functionally differentiating the expression above with respect to g_{ab} we can obtain

$$\langle \hat{T}'^{\ b}_a \rangle_{g'} = \sqrt{\frac{g}{g'}} \langle \hat{T}_a{}^b \rangle_g - \frac{2}{\sqrt{-g'}} g'^{bc} \frac{\delta}{\delta g'^{ac}} \int \mathrm{d}^4 x' \sqrt{-g'} \langle \hat{T}'^{\ c}_c \rangle_{g'} \Omega^{-1}(x') \delta\Omega(x'), \tag{4.67}$$

where we have used the fact that $g'^{ab} \frac{\delta}{\delta g'^{bc}} = g^{ab} \frac{\delta}{\delta g^{bc}}$. We use the subscripts g' and g to denote that the expectation value of the energy-momentum tensor is computed with the vacuum state associated with the g'_{ab} or g_{ab} space-time, respectively. Inserting equation (4.63) above allow us to do the integral, resulting in

$$\langle \hat{T}'^{\ b}_a \rangle_{g'} = \sqrt{\frac{g}{g'}} \langle \hat{T}_a{}^b \rangle_g - \frac{2}{\sqrt{-g'}} g'^{bc} \frac{\delta W'}{\delta g'^{ac}} + \frac{2}{\sqrt{-g'}} g^{bc} \frac{\delta W}{\delta g^{ac}}. \tag{4.68}$$

Substituting for W and W' the two-dimensional analogue of (4.60) and using the conformal properties of the Ricci tensor and Ricci scalar (see, e.g., [4]) reduces eq. (4.68) to

$$\begin{aligned} \langle \hat{T}'^{\ b}_a \rangle_{g'} = \left(\sqrt{\frac{g}{g'}} \right) \langle \hat{T}_a{}^b \rangle_g + \frac{1}{12\pi} \left(\Omega^{-3} \nabla_c \nabla_a \Omega - 2\Omega^{-4} \nabla_c \Omega \nabla_a \Omega \right) g^{cb} \\ + \frac{1}{12\pi} \delta_a{}^b g^{cd} \left(\frac{3}{2} \Omega^{-4} \nabla_c \Omega \nabla_d \Omega - \Omega^{-3} \nabla_c \nabla_d \Omega \right). \end{aligned} \tag{4.69}$$

Now, due to the aforementioned fact that all 2D space-times are conformally flat, we can write the line element of any such space-time in null coordinates (given by eqs. (2.3) and (2.4)) as

$$\mathrm{d}s^2 = \Omega^2(u, v)\ \mathrm{d}u\ \mathrm{d}v. \tag{4.70}$$

Applying (4.69) to this case will give us the renormalized energy-momentum tensor of a massless conformal scalar field in a 2D space-time (here we have made $g' \to g$, with no risk of confusion)

$$\langle \hat{T}_{\mu\nu}(g_{\alpha\beta}) \rangle = \frac{1}{\sqrt{-g}} \langle \hat{T}_{\mu\nu}(\eta_{\alpha\beta}) \rangle + t_{\mu\nu} - \frac{1}{48\pi} R g_{\mu\nu}, \tag{4.71}$$

where

$$t_{\mu\nu} = \begin{pmatrix} -\frac{1}{12\pi}\Omega\partial_u^2\left(\Omega^{-1}\right) & 0 \\ 0 & -\frac{1}{12\pi}\Omega\partial_v^2\left(\Omega^{-1}\right) \end{pmatrix}. \tag{4.72}$$

Equations (4.71) and (4.72) are the fundamental results of this section. They relate the expectation value of the energy-momentum tensor in two conformally related vacuum states. This result will be applied both in the study of accelerated mirrors and of the Hawking flux. We proceed with the former in the next section and the latter in the next chapter.

4.3 Accelerated mirrors as black holes analogues

As we have noted in chapter 1, quantum field theory in curved space-times has also given us insights into flat space-time quantum field theory. One of the most mentioned aspects of this interplay is the Unruh effect, which states that the vacuum state of inertial observers is seen as a thermal state by uniformly accelerated observers. Another aspect of this interplay can be seen when we note that by the Einstein equivalence principle uniform accelerations cannot be distinguished from uniform gravitational fields. This opens the possibility that the study of uniformly accelerated objects in Minkowski space-time can be used to better understand gravity (at least in certain regimes). This turns out to be the case when we consider *accelerated mirrors*. Mirrors are interesting since they provide boundary conditions for the field modes depending on their trajectories. Since these trajectories can have different asymptotic properties they will define different asymptotic positive frequency modes and consequently vacuum states in the past and future may differ. Therefore, *accelerated mirrors can give rise to particle creation*, a fact known as dynamical Casimir effect. Accelerated mirrors are a great laboratory for a first application of the tools developed up to this point since they show many of the effects we will see in the more complicated black hole case, but in a simpler setting.

Our purpose is two-fold: firstly, to comment on how Bogolubov coefficients would be obtained in this case since it is more similar to black holes than the calculations made at the first session of this chapter. Secondly, to lay ground for a model of black hole evaporation presented later.

We postpone an explicit calculation to the next chapter where we treat real black holes. Secondly, we will look at the energy-momentum tensor and later some ideas connecting this with entanglement entropy to study information loss.

Consider a 2D flat space-time, a massless scalar field propagating in such space-time and the field initially in the $|0\rangle_{in}$ state. In null coordinates (eqs. (2.3) and (2.4)), the massless Klein-Gordon equation reduces to

$$\frac{\partial^2 \phi}{\partial u \, \partial v} = 0. \tag{4.73}$$

The presence of a mirror with world-line described by $z = z(t)$ can be imposed by the boundary condition $\phi(t, z(t)) = 0$. Positive frequency modes in the past will be proportional to $e^{-i\omega v}$, evolving into the future into some function of u, which we can write as $e^{-i\omega g(u)}$. Continuity at the surface of the mirror tells us that

$$v = g(u). \tag{4.74}$$

To find the form of $g(u)$ consider that for each u there is a time $t_u = t_u(u)$ such that

$$v = t_u + z(t_u), \tag{4.75}$$

$$u = t_u - z(t_u). \tag{4.76}$$

This can be pictured as the time t_u that a $u = const$ ray propagating from $\mathcal{J}^-$ hits the mirror and leaves as a $v = const$ ray to $\mathcal{J}^+$, as in figure 4. Summing the equations above gives us

$$v = g(u) = 2t_u - u. \tag{4.77}$$

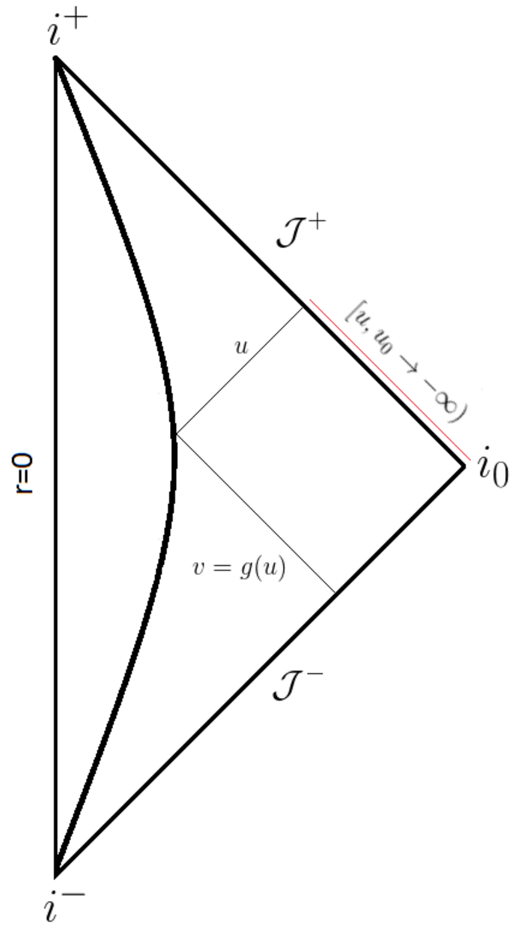

Figure 4: Carter-Penrose diagram of Minkowski space-time with a moving mirror. The highlighted interval in $\mathcal{J}^+$ will be relevant in chapter 6.

To satisfy the boundary condition at the mirror we write the normal modes of $\phi(u, v)$ as

$$f_\omega = \frac{i}{\sqrt{4\pi\omega}} \left(e^{-i\omega v} - e^{-i\omega(2t_u - u)} \right) . \tag{4.78}$$

A given ray u will hit the mirror at different times t_u for different mirror trajectories, therefore changing $g(u)$. Expressing the modes f_ω in terms of positive frequency modes at the future, $F_\omega = e^{-i\omega u}$, allows us to compute the Bogolubov coefficients. Special interest has gone into studying trajectories that are asymptotically null. These can be shown to give thermal emission [3]. Here we have only emphasized the crucial role of determining $g(u)$, since its form will determine the spectrum of the emitted radiation.

To calculate the energy flux radiated from the mirror we first go to a coordinate system where the mirror is at rest. If the trajectory is given by $z = z(t)$, consider the coordinate change

$$u = f(u'), \tag{4.79}$$

$$v = v', \tag{4.80}$$

and define

$$t' = \frac{1}{2}(v' + u'), \tag{4.81}$$

$$z' = \frac{1}{2}(v' - u'). \tag{4.82}$$

The relation $z = z(t)$ can be written, in terms of the null coordinates u' and v', as

$$\frac{1}{2}\left(v' - f(u')\right) = z\left[\frac{1}{2}\left(v' + f(u')\right)\right], \tag{4.83}$$

Imposing that $z' = 0$ (the mirror remains static) means that $f(u')$ satisfies the restriction

$$\frac{1}{2}\left(t' - f(t')\right) = z\left[\frac{1}{2}\left(t' + f(t')\right)\right], \tag{4.84}$$

With this, the flat space-time metric takes the form

$$\mathrm{d}s^2 = (\partial_{u'} f(u'))\mathrm{d}u'\mathrm{d}v'. \tag{4.85}$$

Note that in the (u', v') coordinate system, modes (4.78) can be written as

$$f_\omega = \frac{i}{\sqrt{4\pi\omega}}\left(e^{-i\omega v'} - e^{-i\omega u'}\right) = \frac{1}{\sqrt{\pi\omega}}\sin\left(\omega z'\right)e^{-i\omega t'}. \tag{4.86}$$

These modes are positive frequency modes with respect to $i\partial_{t'}$.

They are the modes that define the static mirror vacuum state. Therefore, we see that the moving mirror is conformally related to the static one (i.e., a "conformal" transformation of the metric relates them). This means that the vacuum states of both situations are also conformally related, allowing us to apply equation (4.71) to this situation. Since a static mirror does not emit radiation, the first term of the right-hand side vanishes and this gives us

$$\langle \hat{T}_{u'u'} \rangle = -\frac{1}{12\pi}\sqrt{\partial_{u'} f(u')}\partial^2_{u'}\left(\frac{1}{\sqrt{\partial_{u'} f(u')}}\right), \tag{4.87}$$

$$\langle \hat{T}_{u'v'} \rangle = \langle \hat{T}_{v'u'} \rangle = \langle \hat{T}_{v'v'} \rangle = 0. \tag{4.88}$$

The state on which the flux is calculated has no particles in the asymptotic past $\mathscr{J}^-$. This boundary condition appears in the fact that there is no energy flux coming in the direction of the mirror ($\langle \hat{T}_{vv} \rangle = 0$), but only going away from the mirror ($\langle \hat{T}_{uu} \rangle \neq 0$). This asymmetry is built in the fact that f is a function of u' only (and not of v'). Due to this fact, we note that the component $\hat{T}_{vv}$ remains unchanged between the two conformally related vacuum states (zero in both cases). We will use this same asymmetry as a shortcut to calculate the energy flux of Hawking radiation in the next chapter. Finally, combining the concepts here with those of chapter 3, a connection can be made between the moving mirror and entanglement entropy of the field. We postpone this discussion to chapter 6, where it will fit more naturally.

5 Black hole thermodynamics and Hawking radiation

In this chapter we show how the first and second law of black hole thermodynamics arise and show the existence of Hawking radiation by combining all tools we presented in the previous chapters. The first law of black hole thermodynamics together with energy conservation lead us to the conclusion that the black hole's mass must decrease and eventually vanish. It follows from this the fact that information is lost during black hole evaporation, a point which we discuss in detail in the next chapter.

5.1 The "laws of black hole mechanics"

Classically, a black hole is a perfect absorber. Due to the causal structure described in chapter 2, everything that falls beyond its horizon encounters the singularity in finite proper time. An interesting question posed initially by Wheeler concerns whether a black hole could be used to violate the second law of thermodynamics, namely, that in closed systems the entropy grows or stays the same. To see why such a question appears, consider a box of gas being thrown into the black hole. Initially there is a non-zero entropy associated with the gas, but after the box "hits" the singularity all the gas is gone and along with it its entropy, therefore violating the second law.

This led Bekenstein to propose that a black hole should, somehow, also possess some entropy [14]. The process of throwing something at it must change this entropy in such a way that the total entropy of the system would satisfy the second law of thermodynamics. This is known today as the *generalized second law of thermodynamics* [15] and it has survived to all *gedanken experiments* proposed so far.

The question was what property of black holes could be associated with its entropy. A natural candidate was the area of the black hole since by this time Hawking had proved the so-called *area theorem* which can be stated as follows:

- *Hawking's area theorem*: Suppose that Einstein's equations are valid and the energy-momentum tensor satisfies the null-energy condition (as described in chapter 2). If space-time contains a black hole with horizon H that is *strongly asymptotically predictable*[23], then the area of H can never decrease.

In order to understand the area-entropy analogy, consider a spherically symmetric[24] black hole with mass M. Its area is given by $A = 16\pi M^2$. If we add an infinitesimal quantity of mass, the area changes by

$$\delta A = 16\pi\delta(M^2) = 32\pi M\delta M, \tag{5.1}$$

which can be rewritten as

$$\delta M = \frac{\delta A}{32\pi M}. \tag{5.2}$$

Since mass can be identified with the black hole energy, this result bears a striking resemblance with the first law of thermodynamics

$$\delta E = T\ \delta S, \tag{5.3}$$

if A is proportional to S. With this argument it is not possible to say what should be the proportionality constant between A and S (which also fixes T). However, other issues were raised about this proposal. Firstly, the laws of black hole mechanics (laid out in [16]) were theorems of differential geometry which should not have any connection to thermodynamics. Secondly, and most importantly, this analogy would imply a non-zero temperature for black holes, something that classically makes no sense. Both issues are only solved when we take quantum mechanics into account, showing us that quantum effects make the black hole radiate thermally. This is the task to which we turn now.

[23] We remind the reader that a black hole is said to be strongly asymptotically predictable if there is a globally hyperbolic region S such that $S \supset I^-(\mathcal{J}^+) \cup \mathcal{H}$ [15].

[24] The theory of black hole thermodynamics can be developed for a much broader class of black holes, including rotating ones. For technical simplicity we will restrict ourselves to spherically symmetric black holes.

5.2 Hawking radiation

We are in position now to derive the existence of Hawking radiation. To do so, consider a massless scalar quantum field $\hat{\phi}$, initially in the state $|0\rangle_{in}$, in a situation in which a spherically symmetric star collapses to form a black hole. At this point we will not be concerned with the black hole evaporation and therefore we will completely disregard all back-reaction effects of the field into the space-time geometry. The Klein-Gordon equation for ϕ in Schwarzschild space-time (whose length element is given by (2.8)) in spherical coordinates is given by

$$\begin{aligned}\nabla^\mu \nabla_\mu \phi = \frac{1}{\sqrt{-g}} \partial_\mu \left(\sqrt{-g}\; g^{\mu\nu} \partial_\nu \phi \right) = \\ \left(1 - \frac{2M}{r}\right)^{-1} \frac{\partial^2 \phi}{\partial t^2} - \frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \left(1 - \frac{2M}{r}\right) \frac{\partial \phi}{\partial r} \right) \\ - \frac{1}{r^2 \sin\theta} \frac{\partial}{\partial \theta} \left(\sin\theta \frac{\partial \phi}{\partial \theta} \right) - \frac{1}{r^2 \sin^2\theta} \frac{\partial^2 \phi}{\partial \varphi^2} = 0.\end{aligned} \tag{5.4}$$

By virtue of time and spherical symmetry of the metric we use as an *ansatz* for the normal modes the form (valid outside the star)

$$f_{\omega l m} \propto \frac{1}{\sqrt{\omega}} Y_{lm}(\theta, \varphi)\, e^{-i\omega t} \frac{g(r)}{r}, \tag{5.5}$$

where $Y_{lm}(\theta, \varphi)$ are the usual spherical harmonics, reducing the wave equation to an ordinary differential equation for $g(r)$:

$$\begin{aligned}\omega^2 \left(1 - \frac{2M}{r}\right)^{-1} g(r) + \frac{2}{r} \left(1 - \frac{2M}{r}\right) \left(\partial_r g(r) - \frac{g(r)}{r} \right) \\ + \frac{2M}{r^2} \left(\partial_r g(r) - \frac{g(r)}{r} \right) + \left(1 - \frac{2M}{r}\right) \left(\frac{2g(r)}{r^2} - \frac{2\partial_r g(r)}{r} + \partial_r^2 g(r) \right) \\ - \frac{l(l+1) g(r)}{r^2} = 0.\end{aligned} \tag{5.6}$$

This equation has not a closed form solution. Therefore, as previously outlined, we will study its asymptotic limits. Firstly, we take $t \to -\infty$ and $r \to \infty$, reducing the preceding equation to

$$\left(\partial_r^2 + \omega^2\right) g(r) = 0, \tag{5.7}$$

whose solution[25],$g(r) = e^{-i\omega r}$, gives us the asymptotic form of the past normal modes $f_{\omega lm}$

$$f_{\omega lm} \propto \frac{1}{\sqrt{\omega}} Y_{lm}\left(\theta, \varphi\right) \frac{e^{-i\omega(t+r)}}{r} = \frac{1}{\sqrt{\omega}} Y_{lm}\left(\theta, \varphi\right) \frac{e^{-i\omega v}}{r}, \tag{5.8}$$

where v and u (appearing below) are the light-cone coordinates of Schwarzschild space-time defined in chapter 2 (here in the appropriate asymptotic limit $r \to \infty$). Exactly the same reasoning with respect to the infinite future (i.e., $t \to +\infty$) gives us the asymptotic form of the future normal modes

$$F_{\omega lm} \propto \frac{1}{\sqrt{\omega}} Y_{lm}\left(\theta, \varphi\right) \frac{e^{-i\omega(t-r)}}{r} = \frac{1}{\sqrt{\omega}} Y_{lm}\left(\theta, \varphi\right) \frac{e^{-i\omega u}}{r}. \tag{5.9}$$

Now our task is to obtain the Bogolubuv coefficients that relate the past and future modes, similarly to what we have done in chapter 4. To obtain these coefficients, we begin by studying the behavior of a mode of the form given by (5.9) as we go from $\mathcal{J}^+$ to $\mathcal{J}^-$. We can write the future modes as a function of the past modes in $\mathcal{J}^-$ as

$$F_{\omega lm} \propto \frac{1}{\sqrt{\omega}} Y_{lm}\left(\theta, \varphi\right) \frac{e^{-i\omega u(v)}}{r}. \tag{5.10}$$

The form of $u(v)$ will determine the Bogolubov coefficients and consequently the mean number of particles created with quantum numbers $\{\omega lm\}$. To specify this function we present an argument based on [12]. Since this argument has some subtleties, an alternative argument is also presented, dealing with a similar (but less realistic) case, in Appendix A.

[25]The choice $-r$ instead of $+r$ in the exponent comes from the fact that we are interested in modes propagating from $r \gg 2M$ to $r = 0$ and therefore we must exclude $+r$, which propagates towards the opposite direction.

We will work here in the geometrical-optics approximation. This means that we will consider modes of a frequency high enough such that we can ignore both interactions between the modes and the star interior and gravitational scattering. It can be shown that the thermal character obtained for the radiation is independent of this assumption, acquiring only extra greybody factors when these assumptions are relaxed. In this limit, let us consider modes of $\hat{\phi}$ that propagate along radial null-geodesics (i.e., θ and φ constants) of Schwarzschild space-time. Using the fact that ∂_t and ∂_φ are Killing vectors of this space-time and that the four-velocity k^μ of this geodesic has zero norm we can write the geodesic equations as

$$\left(1-\frac{2M}{r}\right)\frac{\mathrm{d}t}{\mathrm{d}\lambda} = E, \tag{5.11}$$

$$r^2\frac{\mathrm{d}\phi}{\mathrm{d}\lambda} = L, \tag{5.12}$$

$$\left(\frac{\mathrm{d}r}{\mathrm{d}\lambda}\right)^2 + \frac{L^2}{r^2}\left(1-\frac{2M}{r}\right) = E^2, \tag{5.13}$$

where the geodesic is parameterized by an affine parameter $\lambda \in \mathbb{R}$. We also have made use of conserved quantities associated with both Killing vectors, the energy E and the angular momentum L. Since $\varphi = const$ we have $L = 0$, which leads us to

$$\frac{\mathrm{d}t}{\mathrm{d}\lambda} \mp \left(1-\frac{2M}{r}\right)^{-1}\frac{\mathrm{d}r}{\mathrm{d}\lambda} = 0. \tag{5.14}$$

Using the definition of $r^* = r + 2M\log\left(\frac{r}{2M}-1\right)$, i.e., the tortoise radius, allows us to write the equation above as

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}\left(t \mp r^*\right) = 0, \tag{5.15}$$

as expected from our discussion in chapter 2. $u = const$ characterize the *out-going* radial geodesics and $v = const$ characterize the *in-going* ones.

Taking a specific geodesic $v = v_1 = const$ that crosses the event horizon we can find $u = u(\lambda)$ along this same geodesic. To this end we take the λ-derivative of u

$$\frac{\mathrm{d}u}{\mathrm{d}\lambda} = \frac{\mathrm{d}t}{\mathrm{d}\lambda} - \frac{\mathrm{d}r^*}{\mathrm{d}\lambda} = \frac{\mathrm{d}t}{\mathrm{d}\lambda} - \frac{\mathrm{d}r^*}{\mathrm{d}r}\frac{\mathrm{d}r}{\mathrm{d}\lambda}, \tag{5.16}$$

and use that, according to equations (5.11) and (5.13)

$$L = r^2 \frac{\mathrm{d}\phi}{\mathrm{d}\lambda} = 0, \tag{5.17}$$

$$\frac{\mathrm{d}r}{\mathrm{d}\lambda} = -E, \tag{5.18}$$

obtaining

$$u(\lambda) = 2E\lambda - 4M \log\left(\frac{\lambda}{K_1}\right), \tag{5.19}$$

where we have chosen λ in such a way that $\lambda = 0$ on the event horizon ($r = 2M, u \to +\infty$) and $\lambda < 0$ corresponds to the region $r > 2M$, which implies $K_1 < 0$. Near the event horizon ($\lambda \approx 0$) we have

$$u(\lambda) \approx -4M \log\left(\frac{\lambda}{K_1}\right). \tag{5.20}$$

Now, we wish to find a relation between constant values of v which describe in-going geodesics, originating in $\mathcal{J}^-$, almost entering the horizon, and constant values of $u = u(v) = const$ which describe the resulting out-going geodesics escaping to $\mathcal{J}^+$. To help visualization, consider the Carter-Penrose diagram of the collapse, figure 5.

Analyzing the causal structure above it is possible to see that there exists a (limiting) last in-going geodesic $v = v_0$ that does not cross the horizon. This geodesic characterizes the generators of the horizon. Now, consider a nearby in-going geodesic v that escapes to $\mathcal{J}^+$ as $u = u(v) = const$. We proceed by taking the following steps:

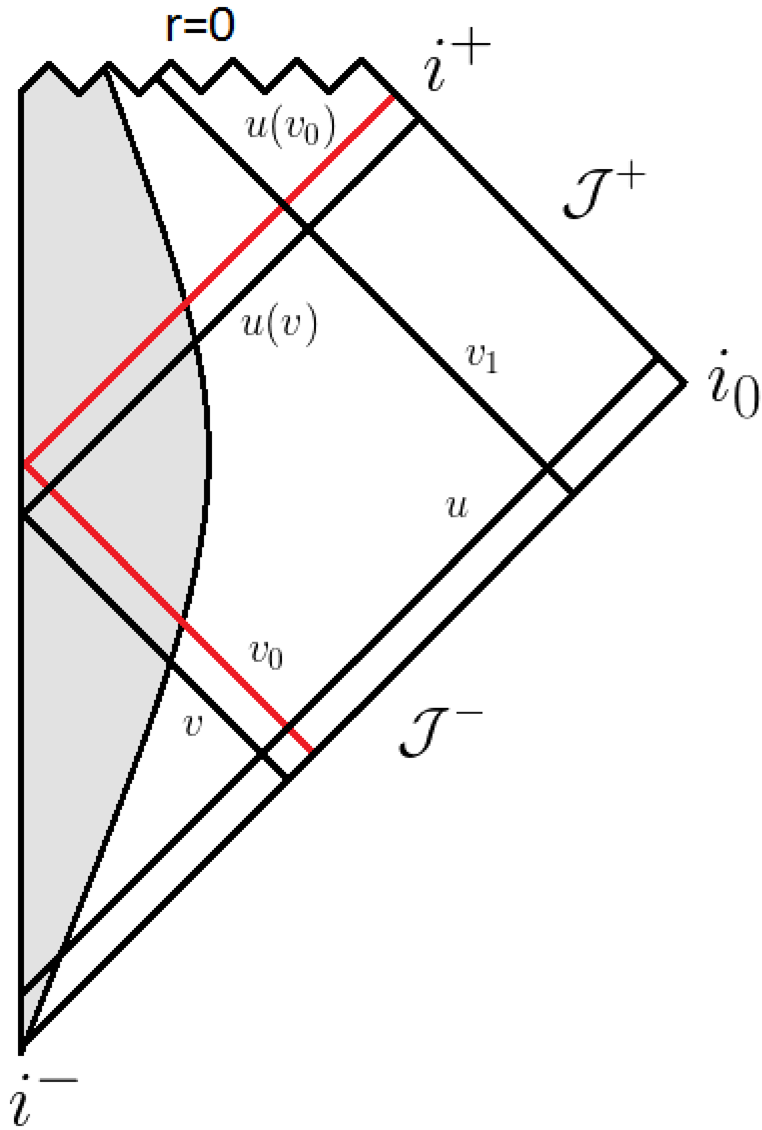

Figure 5: Null-geodesics structure of a collapsing star space-time.

- The set of points characterizing the geodesic v_1 in the Carter-Penrose diagram is described by $v = v_1 = const$ and $u = u(\lambda)$, where λ is the affine parameter parameterizing this geodesic. If we take any other in-going geodesic that crosses the horizon, we can rescale its affine parameter such that the same relation (5.20) holds along all in-going geodesics that cross the horizon.

- Since the out-going geodesic $u(v)$ is characterized by $u = const$, there exists a particular value of λ, call it λ_v, such that equation (5.20) is satisfied for this constant $u(v)$ value, i.e.

$$u(v) = const = -4M \log\left(\frac{\lambda_v}{K_1}\right). \tag{5.21}$$

The same can be said for the limiting out-going geodesic, $u(v_0) \to +\infty$, for which $\lambda_{v_0} = 0$. The affine separation, defined as $\lambda_v - \lambda_{v_0}$, between both out-going geodesics is then simply λ_v. This affine separation, λ_v, is the same along any in-going geodesic that crosses the horizon, due to our choice of scale for the affine parametrization as previously outlined.

- Along the out-going $u = const$ geodesics, the coordinate v is now a function of an affine parameter, i.e., $v = v(\mu)$. Let μ and μ_0 be the values of the affine parameter when the $u = const$ geodesic crosses the in-going $v = const$ and $v_0 = const$ geodesics, respectively. Again, we can choose such parametrization ensuring that the affine separation, $\mu - \mu_0$, is the same along any out-going geodesic.

- One last rescaling makes $\mu - \mu_0 = \lambda_v$. Therefore, the affine separation between the in-going $v = const$ and $v_0 = const$ geodesics is preserved when they turn to out-going $u(v)$ and $u(v_0)$ geodesics.

- Taking the asymptotic limit to $\mathcal{J}^-$, which is the $u = const \to -\infty$ geodesic[26], of (5.11) and (5.13) (with μ as the affine parameter instead of λ) shows us that v is an affine function of μ in this region, i.e.

$$t \approx E\mu + const, \tag{5.22}$$

$$r \approx E\mu + const, \tag{5.23}$$

[26] Equivalently, $r \to \infty$ and $t \to -\infty$.

which means that

$$v \approx t + r \approx 2E\mu + const. \tag{5.24}$$

Therefore, $\lambda_v = \mu - \mu_0 = (2E)^{-1}(v - v_0)$, which we can write as $v_0 - v = K_2\lambda_v$, where, since $v_0 > v$ and $\lambda_v < 0$, $K_2 = const < 0$.

Substituting the relation obtained above into (5.20) we obtain the desired form of $u(v)$:

$$u(v) = -4M \log\left(\frac{v_0 - v}{K_1 K_2}\right) = -4M \log\left(\frac{v_0 - v}{K}\right), \tag{5.25}$$

where $K = K_1 K_2 > 0$. From the Carter-Penrose diagram, it is also clear that modes with $v > v_0$ hit the singularity and do not reach $\mathcal{J}^+$. Therefore, the future modes $F_{\omega lm}$ given by (5.9) when traced back to $\mathcal{J}^-$ can be written as

$$F_{\omega lm} \propto \frac{1}{\sqrt{\omega}} Y_{lm}(\theta, \varphi) \frac{1}{r} \exp\left[4Mi\omega \log\left(\frac{v_0 - v}{K}\right)\right] \theta\left(v_0 - v\right). \tag{5.26}$$

To obtain the Bogolubov coefficients we must expand $F_{\omega lm}$ in terms of $f_{\omega lm}$. Due to spherical symmetry, only modes with the same l and m are coupled. Therefore, we focus on the ω part of the expansion. Applying (4.18) of chapter 4 gives us

$$F_{\omega lm} = \int_0^{+\infty} \mathrm{d}\omega' \left(\alpha_{\omega\omega' lm} f_{\omega' lm} + \beta_{\omega\omega' lm} f^*_{\omega' lm}\right). \tag{5.27}$$

Multiplying both sides above by $e^{\pm i\omega'' v}$ and integrating in v lead us to

$$\int_{-\infty}^{+\infty} \mathrm{d}v \ \exp\left(\pm i\omega'' v\right) F_{\omega lm} = \int_{-\infty}^{+\infty} \mathrm{d}v \ \exp\left(\pm i\omega'' v\right) \int_0^{+\infty} \mathrm{d}\omega' \left(\alpha_{\omega\omega' lm} f_{\omega' lm} + \beta_{\omega\omega' lm} f^*_{\omega' lm}\right). \tag{5.28}$$

Substituting the expressions for the modes at $\mathcal{J}^-$, eqs. (5.8) and (5.26), reduces the equation above to

$$
\begin{aligned}
&\int_{-\infty}^{+\infty} \mathrm{d}v \; \frac{\theta(v_0 - v)}{r\sqrt{\omega}} \exp\Big(\pm i\omega^{''} v\Big) e^{-i\omega u(v)} \\
&\propto \int_0^{+\infty} \mathrm{d}\omega^{'} \frac{1}{r\sqrt{\omega^{'}}} \int_{-\infty}^{+\infty} \mathrm{d}v \; \alpha_{\omega\omega' lm} \exp\Big(-i(\omega^{'} \mp \omega^{''})v\Big) + \\
&\beta_{\omega\omega' lm} \exp\Big(i(\omega^{'} \pm \omega^{''})v\Big),
\end{aligned}
\tag{5.29}
$$

which allows us to obtain $\alpha_{\omega\omega' lm}$ (by choosing the + sign in the exponential) and $\beta_{\omega\omega' lm}$ (by choosing the $-$ sign), given by (after relabeling $\omega'' \to \omega'$)

$$
\alpha_{\omega\omega' lm} \propto \int_{-\infty}^{v_0} \mathrm{d}v \left(\frac{\omega'}{\omega}\right)^{\frac{1}{2}} e^{i\omega' v} \exp\left(4Mi\omega \ln\left(\frac{v_0 - v}{K}\right)\right), \tag{5.30}
$$

$$
\beta_{\omega\omega' lm} \propto \int_{-\infty}^{v_0} \mathrm{d}v \left(\frac{\omega'}{\omega}\right)^{\frac{1}{2}} e^{-i\omega' v} \exp\left(4Mi\omega \ln\left(\frac{v_0 - v}{K}\right)\right). \tag{5.31}
$$

where we have omitted the same proportionality constant in both equations. Firstly, we focus on the $\alpha_{\omega\omega' lm}$ integral. Changing variables from v to $s = v_0 - v$ transforms (5.30) into

$$
\alpha_{\omega\omega' lm} \propto \int_0^{\infty} \mathrm{d}s \left(\frac{\omega'}{\omega}\right)^{\frac{1}{2}} e^{-i\omega' s} e^{i\omega' v_0} \exp\left(4Mi\omega \ln\left(s/K\right)\right). \tag{5.32}
$$

Since the integrand is analytic in the lower right quadrant of the complex plane, we use a quarter-circle contour with radius R, outlined in figure 6, to rewrite the integral. Region II of the contour does not contribute when $R \to \infty$ and therefore

$$
\alpha_{\omega\omega' lm} \propto -\int_{-i\infty}^{0} \mathrm{d}s \left(\frac{\omega'}{\omega}\right)^{\frac{1}{2}} e^{-i\omega' s} e^{i\omega' v_0} \exp\left(4Mi\omega \ln\left(s/K\right)\right). \tag{5.33}
$$

Making $s = is'$ turns it to

$$
\alpha_{\omega\omega' lm} \propto -i\int_{-\infty}^{0} \mathrm{d}s' \left(\frac{\omega'}{\omega}\right)^{\frac{1}{2}} e^{\omega' s'} e^{i\omega' v_0} \exp\left(4Mi\omega \ln\left(is'/K\right)\right). \tag{5.34}
$$

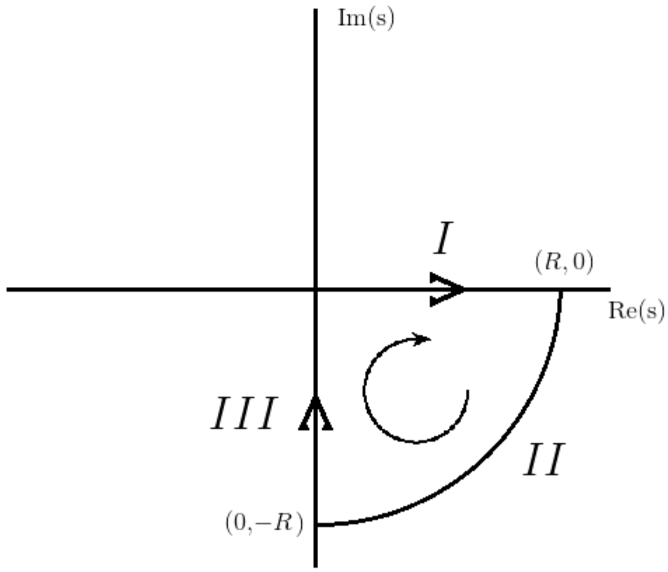

Figure 6: Integration contour used to calculate the integral in (5.30).

Since in this range of integration we have

$$\ln\left(is'/K\right) = \ln\left(-i|s'|/K\right) = \ln\left(|s'|/K\right) - \frac{i\pi}{2}, \tag{5.35}$$

the integral can be written as

$$\alpha_{\omega\omega' lm} \propto -ie^{2\pi\omega M} \int_{-\infty}^{0} \mathrm{d}s' \left(\frac{\omega'}{\omega}\right)^{\frac{1}{2}} e^{\omega' s} e^{i\omega' v_0} \exp\left(4Mi\omega \ln\left(|s'|/K\right)\right). \tag{5.36}$$

A similar idea applied to $\beta_{\omega\omega' lm}$ leads us to

$$\beta_{\omega\omega' lm} \propto ie^{-2\pi\omega M} \int_{-\infty}^{0} \mathrm{d}s' \left(\frac{\omega'}{\omega}\right)^{\frac{1}{2}} e^{\omega' s} e^{-i\omega' v_0} \exp\left(4Mi\omega \ln\left(|s'|/K\right)\right). \tag{5.37}$$

Taking the square of the absolute value of $\alpha_{\omega\omega' lm}$ and $\beta_{\omega\omega' lm}$ shows that they satisfy

$$|\alpha_{\omega\omega' lm}|^2 = e^{8\pi M\omega} |\beta_{\omega\omega' lm}|^2. \tag{5.38}$$

Using equation (4.20) of chapter 4, with an integral representation for the delta distribution we obtain

$$\int \mathrm{d}\omega' \left(|\alpha_{\omega\omega'}|^2 - |\beta_{\omega\omega'}|^2\right) = \lim_{T\to\infty} \lim_{\omega'\to\omega} \frac{1}{2\pi} \int_{-T/2}^{T/2} \mathrm{d}t \; e^{i(\omega-\omega')t}, \tag{5.39}$$

which give us the formal expression

$$\lim_{T\to\infty} \frac{T}{2\pi} = \int \mathrm{d}\omega' \left(|\alpha_{\omega\omega'}|^2 - |\beta_{\omega\omega'}|^2\right). \tag{5.40}$$

Using relation (5.38) finally yields

$$\lim_{T\to\infty} \frac{T}{2\pi} = \left(e^{8\pi M\omega} - 1\right) \int \mathrm{d}\omega' |\beta_{\omega\omega'}|^2. \tag{5.41}$$

Therefore, the mean number of particles per unit of proper observer time for asymptotic observers at $\mathscr{J}^+$ is given by

$$\langle N_\omega \rangle_{\mathscr{J}^+} \mathrm{d}\omega = \frac{1}{2\pi} \frac{\mathrm{d}\omega}{e^{8\pi M\omega} - 1}. \tag{5.42}$$

The expression above is the same as the mean number of particles in radiation emitted by a blackbody at a temperature

$$T_H = \frac{1}{8\pi M}, \tag{5.43}$$

or

$$T_H = \frac{\kappa}{2\pi}, \tag{5.44}$$

where $\kappa = 1/4M$ is the surface gravity of the black hole defined by equation (2.15). An extremely similar calculation done for the Kerr-Newman space-time would give us the same result as (5.44). Noting that, due to the first law of thermodynamics, $T = \frac{\partial M}{\partial S}$ allows us to find S as a function of M. A simple integration then gives

$$S = \frac{k_B c^3 A}{4G\hbar}, \tag{5.45}$$

where we have restored dimensional units. The presence of $\hbar$, G and k_B combined shows that there is a deep connection between gravity, quantum mechanics and thermodynamics. As we have alluded earlier, this connection is not fully understood today. Hawking radiation is the final piece needed to make black hole thermodynamics a consistent application of classical thermodynamics to black holes. We also note that the same result for the Hawking temperature has been obtained by a myriad of methods, like with the use of Euclidean path integrals (see, e.g., [3]), showing its robustness. Besides the mean number of particles calculated above, another quantity of interest is the energy flux carried away by the radiation. We turn to this question now, applying the machinery developed in the previous chapter to compute the renormalized energy flux of the Hawking radiation.

5.3 Energy flux of the radiation

With the renormalization procedure outlined in chapter 4 we can study the energy-momentum tensor as measured by observers in the asymptotic future $\mathscr{I}^+$. This will shows us that the radiation carries positive energy away from the black hole which will lead us directly to the conclusion that the black hole must evaporate in a finite time.

Consider the $t - r$ plane of the Schwarzschild metric written in its light-cone coordinates $u - v$

$$\mathrm{d}s^2 = \left(1 - \frac{2M}{r(u,v)}\right) \mathrm{d}u \ \mathrm{d}v = \Omega^2(r) \ \mathrm{d}u \ \mathrm{d}v. \tag{5.46}$$

Just as in the moving mirror case, we can find a coordinate transformation $(u, v) \to (u', v')$ that transforms the normal modes of the 2D wave equation we solved (disregarding θ and ϕ) into normal modes that have the same functional form of (4.86), i.e., proportional to $e^{-i\omega t'}$, where $t' = (1/2)(u' + v')$[27]. This means that the vacuum state defined by these modes will be conformally related to $|0\rangle_{in}$, similarly to the mirror case, making the first term of the right-hand side of (4.71) vanish.

[27] For more details, see [3]

The asymmetry of boundary conditions (no incoming radiation) means that the coordinate transformation changes u only, i.e., $u = f(u')$ and $v = v'$. The resulting metric will have the form

$$\mathrm{d}s^2 = \left(1 - \frac{2M}{r(u,v)}\right)\frac{\mathrm{d}u}{\mathrm{d}u'}\mathrm{d}u'\ \mathrm{d}v' = \Omega^2(r)\frac{\mathrm{d}u}{\mathrm{d}u'}\ \mathrm{d}u'\mathrm{d}v'. \tag{5.47}$$

Consider now the $v'v'$ component of (4.71). We have

$$_{in}\langle 0|\hat{T}_{v'v'}|0\rangle_{in} = {}_{in}\langle 0|\hat{T}_{vv}|0\rangle_{in} = t_{vv} = -\frac{1}{12\pi}\Omega\partial_v^2\Omega^{-1}. \tag{5.48}$$

Since Ω is a function of r only, we can substitute

$$\frac{\partial}{\partial v} = \frac{\partial r^*}{\partial v}\frac{\partial}{\partial r^*} = \frac{\partial r^*}{\partial v}\frac{\partial r}{\partial r^*}\frac{\partial}{\partial r} = \frac{1}{2}\Omega^2(r)\frac{\partial}{\partial r}, \tag{5.49}$$

and obtain

$$_{in}\langle 0|\hat{T}_{vv}|0\rangle_{in} = \frac{1}{24\pi r^4}\left(-Mr + \frac{3}{2}M^2\right). \tag{5.50}$$

Calculating the expression above at the horizon ($r = 2M$) leads to

$$_{in}\langle 0|\hat{T}_{vv}(r = 2M)|0\rangle_{in} = -\frac{1}{768\pi M^2}. \tag{5.51}$$

We see that a negative energy flux enters the horizon. Now, since $_{in}\langle 0|\hat{T}_{ab}|0\rangle_{in}$ is divergenceless, by construction, we can apply Gauss theorem to show that this must correspond to an equal magnitude energy flux at infinity, $_{in}\langle 0|\hat{T}_{uu}(r \to \infty)|0\rangle_{in}$ given by (5.51) with opposite sign (all other terms vanish at $\mathcal{J}^+$). Since the only other source of energy besides radiation is the gravitational field, this means that *the black hole loses mass*. This appears contradictory to the area theorem we stated earlier, but the *quantum* field violates one of the key hypotheses of this theorem, namely that the field satisfies the null energy condition.

In deriving the thermal spectrum of Hawking radiation we have supposed that space-time would not suffer any effect of back-reaction by the radiation.

When asking what is the final fate of the space-time we are describing this clearly cannot be the case. Due to the radiation being (approximately [17]) thermal, we can use the Stefan-Boltzmann law as an order of magnitude approximation to compute the rate of energy loss of the black hole according to asymptotic observers. Either way, we have

$$\frac{\mathrm{d}M}{\mathrm{d}t} \approx -16\pi M^2 T^4 \approx -M^{-2}. \tag{5.52}$$

It is then simple to integrate this equation to obtain

$$M(t)^3 - M_0^3 \approx -(t - t_0) = -\Delta t, \tag{5.53}$$

where M and t_0 are the initial mass of the black hole and the initial time of measurement. Note that, for a fixed (but arbitrarily large) finite radius r, we can write $\Delta u = u - u_0 = t - r^* - t_0 + r^* = t - t_0 = \Delta t$ and we can write M as a function of u. This will be useful in the next chapter. Setting aside for the moment the (important) fact that the semi-classical picture depicted here may change drastically in the final moments of the evaporation, we can see that the black hole will evaporate in a *finite* time given by

$$\Delta t \approx M^3. \tag{5.54}$$

After this, the final configuration is an approximately Minkowski space-time together with thermal radiation. The quantum state of radiation will be a mixed state given by an expression of the type (3.11). This means that our initial state $|0\rangle_{in}$ has evolved to a mixed state. This kind of evolution cannot be achieved by a unitary operator, as proved in chapter 3. Therefore, *information was lost*. This completes our posing of the so-called black hole information loss, which we know discuss in detail.

6 Information loss in black holes

We have seen in the last chapter that the existence of Hawking radiation is, on one hand, necessary to the self consistency of black hole thermodynamics, but on the other hand, leads to information loss in black hole space-time. Several aspects of this fact can be discussed. Firstly, we treat common objections to the Hawking effect. Secondly, the question of whether information loss constitutes a paradox by itself is debated. Finally, common proposals for guaranteeing information retrieval in a (unitary) theory of quantum gravity are commented upon.

6.1 Objections against Hawking radiation

Several points of our earlier derivation can be questioned, including the validity of the Hawking temperature expression (5.43) when space-time is changing, the geometrical optics approximation used and the assumption that the Stefan-Boltzmann law holds up to complete evaporation of the black hole.

First, our calculations presented earlier assume no back-reaction of the radiation energy-momentum tensor on the space-time. This will change the metric and the behavior of geodesics in this space-time, which will lead to different Bogolubov coefficients. A self-consistent evolution of both the space-time and radiation coupled through Einstein semi-classical equations has not been achieved yet, due mainly to difficulties with respect to renormalization of ${}_{in}\langle 0|\hat{T}_{ab}|0\rangle_{in}$ in four-dimensional space-times. Nevertheless, approximate calculations using 2D models and spherical symmetry have showed that the general expression for the flux given by (5.52) is valid as long as $M \gg M_P$, where M_P is the Planck mass [12].

The second objection deals with the geometrical optics approximation. This approximation enters our calculations when we consider geodesics to find $u(v)$ instead of solving completely the Klein-Gordon equation (including the scattering potential term).

This means that we only consider modes with high frequencies (including modes with frequencies above the Planck scale), which, in turn, suffer an arbitrarily large exponential redshift. A way to stay entirely in the semi-classical approximation would be to cut off these ultra-high frequency modes, but this would lead to Lorentz invariance violation. Studies in analogue models of gravity have shown that even when there's breaking of this symmetry the (approximate) thermal character of the radiation survives [18].

The final objection is in regard to the assumption that evaporation proceeds completely. This is crucial assumption and its validity can only be answered when we have a complete theory of quantum gravity. Nevertheless, as we will see soon, ways of preventing complete evaporation introduces concepts that have their own problems, some of which much worse than information loss. In the semi-classical theory there is no compelling reason to suppose that evaporation is halted.

We note, finally, that Hawking radiation has been derived in a number of different ways, some of which include approximations similar to ours and others which start from different hypotheses, like [19], using no Bogolubov coefficients or the geometrical optics approximation.

With these objections addressed, the causal structure of space-time, including the post-evaporation phase, can be represented by the diagram of Figure 7, where we have also drawn two space-like surfaces, Σ_1 and Σ_2.

6.2 Is information loss a paradox?

Let us turn our attention now to answering the main question of this work. As we have argued at the end of the previous chapter, information is indeed lost in black hole evaporation, as initially pointed out in [20]. To see this in the causal structure of space-time, consider the space-like surface Σ_1 in figure 7. This surface is not a Cauchy surface since it can be proven that Σ_2 is not in $D^+(\Sigma_1)$. Naively, this might be telling us that we cannot predict what happens after the evaporation from knowledge of the state of the field in Σ_1.

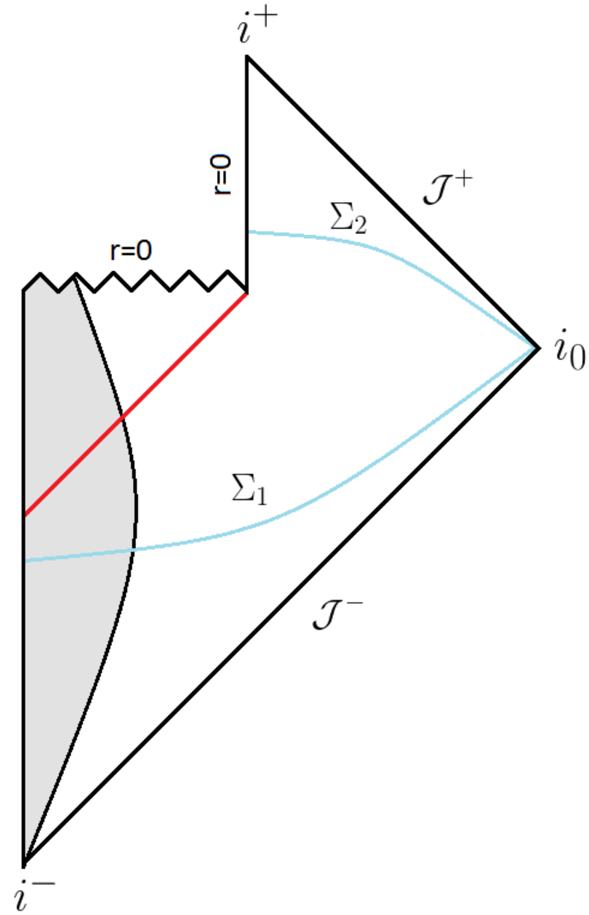

Figure 7: Space-time of a collapsing star forming a black hole which eventually completely evaporates.

However, this need not to be the case. Consider a massless scalar field propagating in Minkowski space-time with a point P removed. The values of the field inside the future light-cone of P appear to be undetermined. But, by continuity, the value of the field at P can be determined, thus allowing us to evolve the field inside the future-light cone of P. Even in this non-globally hyperbolic space-time, the field evolution is determined uniquely from initial data on a space-like surface (such as $t = const$ hypersurface). Such a situation is depicted in the figure 8.

In our evaporating black hole space-time, the role of the point in the example above is taken by the naked singularity at the end of the evaporation. The semi-classical theory then hypotheses that this naked singularity (which exists for an arbitrarily small amount of time)

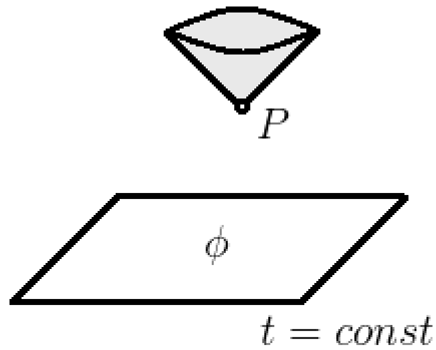

Figure 8: Evolution in Minkowski space-time with one point removed.

does not influence space-time. Therefore, for all practical purposes, Σ_1 "acts" like a Cauchy surface and given initial data on it we can evolve it both to the future and to the past, obtaining the state of the field (i.e., its value and of its conjugate momentum) at any given time as described in chapter 1 for classical fields. The same applies to quantum fields.

Now, consider the space-like surface Σ_2. This surface is also not a Cauchy surface since its domain of dependence $D(\Sigma_2)$ is not the whole space-time. This can be easily seen by considering inextendible causal future curves from any point inside the event horizon before the black hole evaporates, which do not cross Σ_2. This case, however, is different from above. Some degrees of freedom from the field propagate into the singularity and are destroyed, leaving behind only a thermal state after evaporation. Different initial states will give rise to the same final state. Therefore, given data over Σ_2 it is not possible to evolve it backwards in time, although we can still evolve it into the future.

This is a way of seeing how the singularity effectively *opens* the system and render evolution non-unitary. Although information is lost, there is no *paradox*, since information does not need to be preserved in open quantum systems. A similar phenomenon can happen even in flat-space time [21], as we explain below.

Consider 2D Minkowski space-time, $(\mathbb{R}^4, \eta_{ab})$ and a set of hyperbolic space-like surfaces Σ_c of the type $t^2 - z^2 = c^2$. These surfaces are *not* Cauchy surfaces since some light-like causal inextendible curves do not hit them. Therefore, $D(\Sigma_c) \neq M$. If we give initial data in a surface $t = 0$ of this space-time describing a pure state of a field, the state of the field will be mixed when considered over the hyperbolic surfaces. Again, information is lost. A depiction of this situation can be seen in figure 9. The dashed part of future light-like infinity $\mathcal{J}^+$ is what is missing to make Σ_2 a Cauchy surface. We note that since we are dealing with a globally hyperbolic space-time, it is easier to see that information loss in this case does not imply a breakdown of quantum theory.

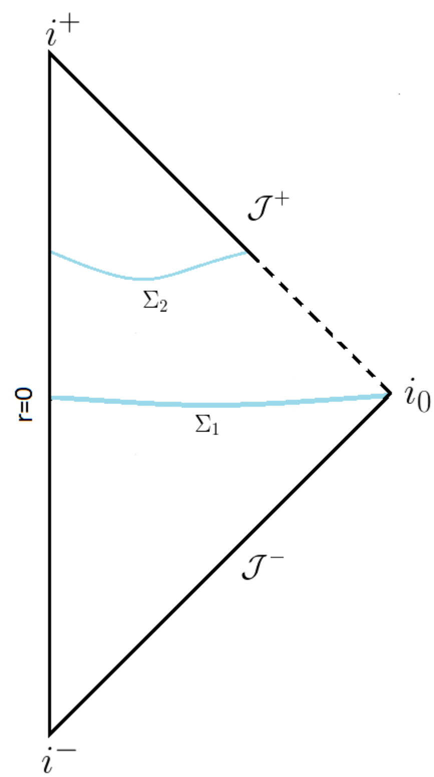

Figure 9: (Non-unitary) Evolution from a space-like surface Σ_1 to another space-like surface Σ_2.

Summarizing our argument, *there is no information loss paradox in the semi-classical theory, despite the fact that information is lost.*

6.3 Imposition of unitarity and the relevance of quantum gravity

As described above, information loss in the semi-classical theory of black holes is a fact and poses no challenge to the theory. Despite this fact, we may still expect that a full theory of quantum gravity (for which semi-classical gravity would be a low-energy approximation) will unveil the real fate of the black hole singularity, eliminating information loss and making evolution completely unitary. If this is the case, we would be able (in principle, not necessarily in practice) to recover all information about the initial quantum state.

To close this chapter we want to comment on some proposed fates of the information in evaporating black hole space-times, using an analog model based on accelerated mirrors to motivate a mathematically more precise way of saying what is the "fate" of information. Some of these proposals are still highly debated today and a vast amount of literature dealing with them exists. A full analysis of each of them vastly transcends the scope of this work. Therefore, we only briefly comment their main points, emphasizing that this question can only be fully solved by a complete theory of quantum gravity. Finally, we note that this list is by no means exhaustible[28] and experimental proof of any of these proposals is still lacking.

6.3.1 Black hole evaporation analogue and the entanglement entropy curve

Here we will use concepts of chapters 3 and 4 together with some machinery of conformal field theory to construct an analogue of evaporating black holes with accelerated mirrors. Consider again a 2D flat space-time, in which a mirror moves in such a way that it imposes relation (4.77) between the null-coordinates u and v. Take the Carter-Penrose diagram of figure 4 of chapter 4. By applying equation (3.14) to the density matrix formed from the in-vacuum $|0\rangle_{in}$ of a massless scalar field when the region $(u, +\infty)$ is traced out, with some modifications to regularize and renormalize the calculation[29], it is possible to define a

[28] One notable omission here is the so-called AdS/CFT correspondence. This is due to the fact that our main focus during this work were asymptotically flat space-times. Most of what we have done here can also be done for Anti-de Sitter space-times, but conclusions coming from the AdS/CFT correspondence regarding information loss seem hard to transpose to asymptotically flat or de Sitter space-times.

[29] In chapter 3 we considered only finite quantum systems. Renormalization is required for quantum fields due to ultra-violet divergences.

renormalized entanglement entropy given by [22]

$$S(u) = -\frac{1}{12\pi} \log\left(\partial_u g(u)\right). \tag{6.1}$$

This formula also assumes that the mirror is asymptotically past inertial ($\partial_u g(u) \to 1$ as $u \to -\infty$)[30]. The curve $S(u) \times u$ is called the *Page curve*[31]. It can be interpreted as the evolution of entanglement entropy as a function of time. Although we defined such curve in Minkowski space-time for the situation of a moving mirror, a similar Page curve could be constructed for the evaporating black hole scenario. *If* in such case all the entropy in space-time after the black hole evaporation is encoded in the radiation entanglement entropy[32] then imposition of complete unitarity amounts to saying that the Page curve goes to 0 as $u \to +\infty$.

The condition $S(u \to +\infty) \to 0$ is *not* fulfilled for the final state of semi-classical black hole evaporation since this state would have a non-zero entropy (in fact, it is given by the entropy of a thermal state). To construct the non-unitary curve, consider equations (5.45) and (5.53) (when expressed in terms of u instead of t), that, combined, give for the Bekenstein-Hawking entropy

$$S_{BH}(u) = \frac{A(u)}{4} = 4\pi M(u)^2 \approx 4\pi \left(M_0^3 - u\right)^{2/3}. \tag{6.2}$$

The entropy of the radiation starts from zero and grow to its maximum value, $\approx 4\pi M_0^2$, when the black hole completely evaporates. The behavior of the Bekenstein-Hawking entropy and of the radiation entanglement entropy in the semi-classical picture is sketched in figure 10. In the next figure, a different curve for the radiation entropy, imposing that the final state is pure, is drawn. Both these curves assume that all entropy in the is system entanglement entropy.

[30]If this was not the case, the state $|0\rangle_{in}$ would not be pure

[31]Strictly speaking, the name *Page curve* only applies when $S(u \to +\infty) \to 0$, but here we will use the name for general $S(u) \times u$ curves.

[32]This basically assumes no hidden quantum-gravitational degrees of freedom.

The question that some proposed "solutions" to the black hole information loss "problem" attempt to solve is *how* this curve arises (and *how* to compute it).

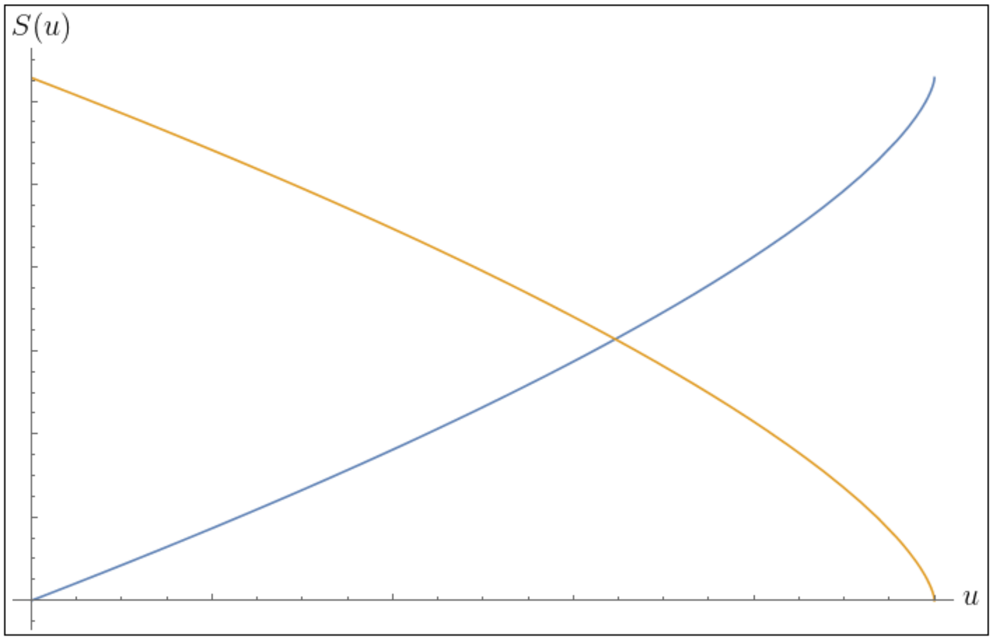

Figure 10: Qualitative sketch of the radiation entanglement entropy (blue) and Bekenstein-Hawking entropy (orange) as a function of time *in the semi-classical picture*.

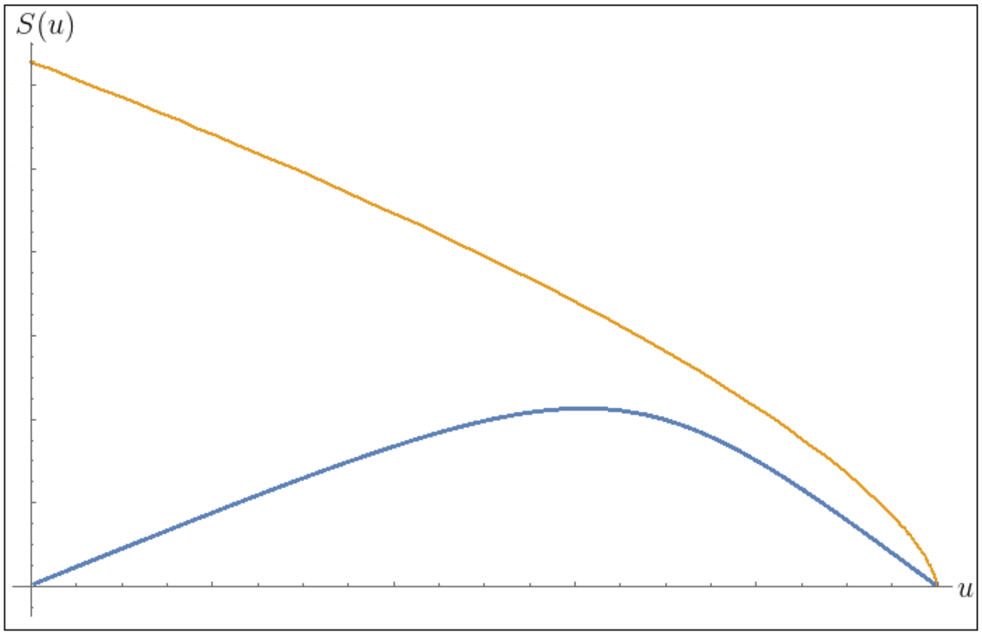

Figure 11: Qualitative sketch of the radiation entanglement entropy (blue) and Bekenstein-Hawking entropy (orange) as a function of time *when unitarity is imposed*. This is the *Page curve*.

6.3.2 Baby universes and connected universes

One way to ensure that semi-classical gravity is not violated while at the same time enforcing unitarity is to consider only effects happening at the Planck scale. Therefore, most ways of preserving information about the initial data imply saying something about the singularity. In some of such models the singularity is substituted by another flat space-time region (i.e., either it gives rise to a "baby universe" or it is connected to another universe) [23]. Naively, information is lost in our universe for all practical purposes, but it is fundamentally preserved since there is no singularity to open the system. The Page curve would be that of figure 10 when restricted to our universe. Two possible Penrose diagrams for these situations are shown in the figures below.

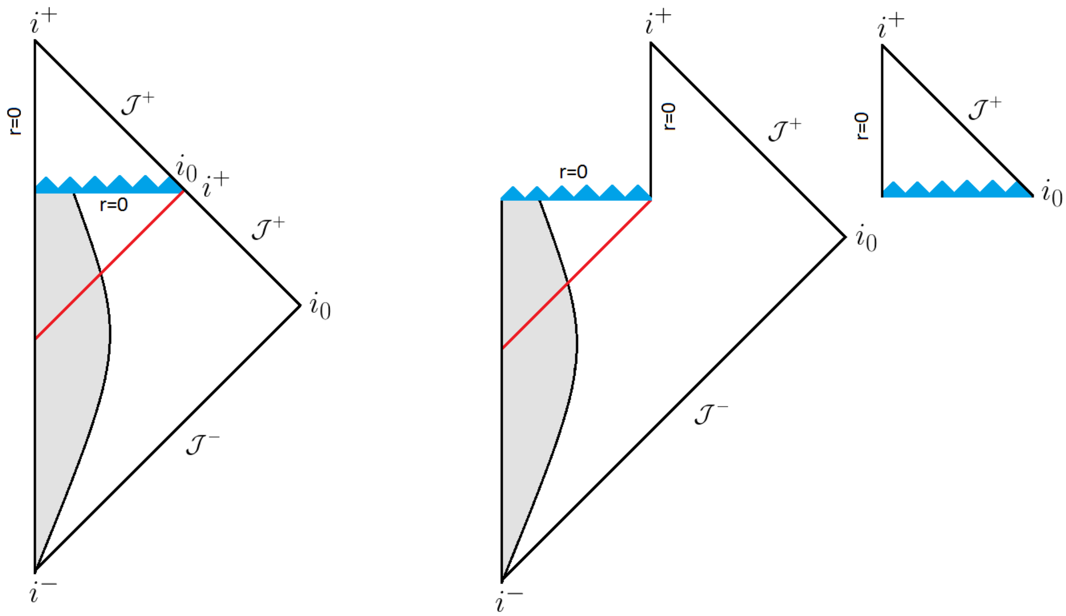

Figure 12: Examples of information retrieval proposals involving different universes. In the left we have a baby universe where quantum gravity effects leave a Planck-sized black hole and in the right connected universes where the quantum gravity regions in blue are identified, but the black hole fully evaporates.

This is an unpopular way of preventing information loss in quantum gravity. Firstly, it requires us to believe in a multiverse. Secondly, in the absence of a quantum-gravitational mechanism that allows this process to occur, we can see that it is very *ad hoc* and unjustified, besides being very hard (if not impossible) to probe experimentally.

6.3.3 Remnants

A second solution along the lines of only modifying the theory at the Planck scale is assuming that the black hole does not evaporate completely, but leaves behind a Planck sized remnant. Information would be trapped in the remnant, inaccessible to external observers, but again, fundamentally preserved. There are some problems with this approach, however (which, some argue, can be sidestepped).

One is that a Planck sized remnant should form regardless of the initial size (and, therefore, initial entropy) of the black hole. This would imply that a fixed sized remnant should be able to hold an arbitrarily large amount of information, which seems improbable. A related problem has to do with amplitudes for common quantum field theory calculations, which should change if remnants exist (since there should be an infinite number of different remnants, even an almost-zero amplitude for remnant formation would be relevant). An interesting aspect of this proposal is that if the number of remnants existing in the universe is too large they could be experimentally probed (for instance, by high-precision gravitational lensing). The Page curve would also be the one of figure 10. No fundamental information loss would exist, but observers without access to the remnant interior would see a mixed state (since they would trace out the remnant interior). A conformal diagram of this situation is shown in the next figure.

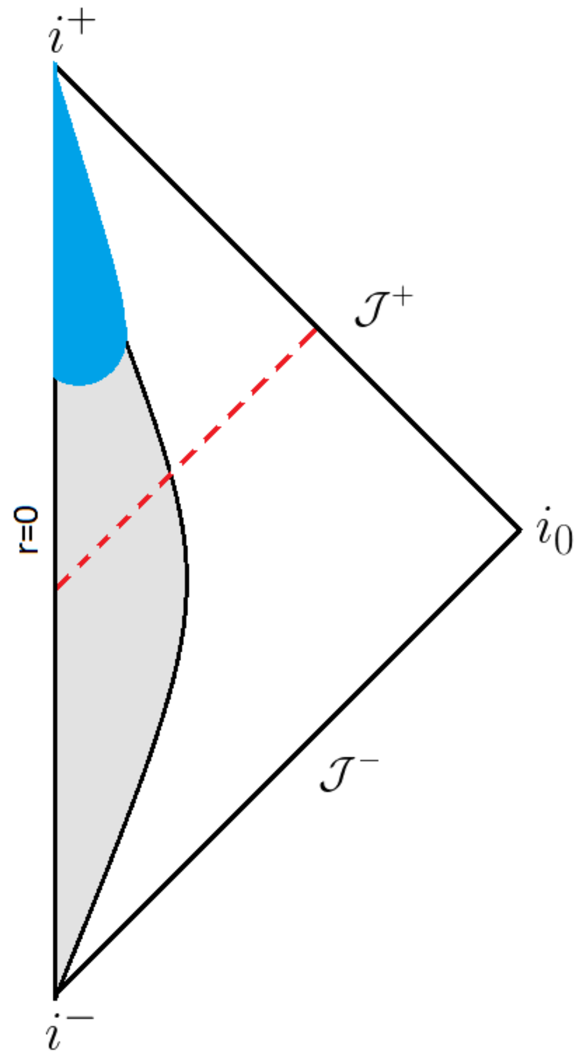

Figure 13: A black hole evaporating and leaving a Planck-sized remnant. The blue region is the one influenced by quantum gravity. The dashed red line is what would have been the black hole horizon (now only an apparent horizon).

6.3.4 Meta-stable black holes

Another way to prevent information loss is if there is nowhere for information to be lost in the first place. This would happen if a true black hole could never form. Since we know that objects with black hole properties exist in our universe, we cannot simply say that black holes do not exist.

However, we could argue that they are only *meta-stable* black holes (i.e., objects that look like real black holes for a long observational time) [24]. This seems to be a very underdeveloped proposal, but if this would be the case (and if there was no singularity), information would be trapped inside of the black hole by a very long time, but would eventually get out and be accessible for external observers again. This is a radical way to prevent information loss since classical general relativity black holes should form in natural scenarios (like stellar collapses). Therefore, this would require a modification of semi-classical gravity at *low* energies.

6.3.5 Black hole complementarity and firewalls

Most proposals for information loss do not only ask for unitarity during the whole spacetime evolution. They also demand non-violation of low-energy physics and the validity of general relativity and QFTCS at large scales, both reasonable claims. Another tacit assumption, made in some theories of quantum gravity, is that Bekenstein-Hawking entropy counts the number of quantum-gravitational microstates of the black hole.

One way to preserve all information lost inside the horizon would be to copy such information in a way that one copy remains outside the horizon. Disregarding the question of what would do this, such copying cannot be made unitarily (it violates the no-cloning theorem). Nevertheless, it has been argued that such copying can occur in a way that no single observer sees a violation of the no-cloning theorem [25]. This is known as *black hole complementarity*. In a certain sense, one part of quantum mechanics (no-cloning) is violated (although not in an experimentally detectable way) to preserve another part of quantum mechanics (unitarity, which, as we saw, is *not* violated). This would force the Page curve to be that of figure 11.

However, even setting this issue aside, it has been recently argued that imposing simultaneously all conditions described above leads us to the conclusion that a freely falling observer should observe the state of the field as a highly energetic one near the horizon. This came to be known as the *firewall* argument [26].

A simple explanation for how it arises goes as follows: the vacuum state of the field is an entangled state when we divide space into what is inside and what is outside of the event horizon (similar to what was described in chapter 3). This is necessary to guarantee regularity at the horizon (i.e., that the energy-momentum tensor does not diverge there). On the other hand, in the semi-classical picture, if information is to be regained after the evaporation, there must be correlation between the "old" Hawking radiation (quanta created long before the black hole evaporates) and "new" Hawking radiation (quanta created near the end of evaporation). It can be shown that these properties cannot be simultaneously satisfied. Not abandoning unitarity then means that the energy-momentum tensor must diverge at the horizon (a "firewall").

In our opinion, the firewall argument is an interesting demonstration that enforcing unitarity in the semi-classical theory cannot work. This again reflects the role that the singularity plays in the question of information loss.

6.3.6 Loop quantum gravity degenerated states

Another way to escape information loss is to suppose that quantum gravity introduces new degrees of freedom at energies higher than the Planck energy (such as not to alter the semi-classical picture). One example of theory that proposes something along these lines is loop quantum gravity [27]. In this theory, space-time is treated as a sort of eigenstate of a certain geometrical operator. As such, Minkowski space-time could be, somehow, degenerated and can arise from multiple different quantum-gravitational microscopic degrees of freedom. If this is the case, different microscopic states with the same macroscopic state (Minkowski space-time) could keep information that would be lost in the semi-classical picture. Therefore, information would be fundamentally preserved, but inaccessible for observers having access to only low energy degrees of freedom. This proposal has its problems, like the others. Among them is that, without access to Planckian energies, it is completely experimentally indistinguishable from information loss.

7 Conclusions and outlook

This work has concerned itself with posing, clarifying and analyzing information loss in black holes. Despite more than forty years of existence, the scenario we have described is still posed as a paradox by some and a violation of quantum field theory in the semi-classical regime.

As we have seen, in the semi-classical approach black holes emit radiation thermally and this leads to a complete different picture when compared to classical general relativity, a picture where the black hole evaporates completely and leaves behind nothing but flat space-time (plus radiation). Information present after the evaporation is not enough to reconstruct the entire evolution of the field in the resulting space-time. This is reflected in the fact that space-like surfaces after the evaporation do not include the interior of the black hole prior to the evaporation in their past domains of dependence. Therefore, degrees of freedom that fall into the singularity leave no imprint on them. As pointed out in chapter 2, when this is the case even classical field theory loses its power to predict the past state of the field.

For the sake of complete clarity we state one last time our conclusion of chapter 6. *There is no black hole information loss paradox in the semi-classical theory.* Unitarity cannot be maintained throughout the whole space-time due to the existence of the singularity, which effectively makes the space-time/quantum field system "open", rendering the fundamental axiom about evolution in quantum mechanics *not* applicable to this case. Also, we have showed that this can happen even in flat space-time, where no one doubts the validity of either special-relativity or quantum theory.

Although there is no paradox, it is (mostly) agreed that information is indeed lost in the (semi-classical) process of black hole evaporation. The fate of this information remains unknown and we believe this will be the case as long as we do not have a complete theory of quantum gravity that can make predictions in the strong curvature regime. In the last chapter we have touched upon this subject by showing recent proposals for the fate of such information that are still being debated nowadays, citing only a small sample of a cornucopia of different ideas, ranging from hidden degrees of freedom in quantum gravity to meta-stable black holes and beyond.

All these solutions propose to "solve the black hole information problem" but turn out to create either drastic violations of QFTCS in the semi-classical regime (e.g., firewalls) or violations of classical general relativity by some unknown means (e.g., meta-stable black holes). Others are almost impossible to probe experimentally and do not differ much from true information loss (e.g., baby universes).

Our opinion is that the problem with most proposed ways to impose information retrieval from evaporating black holes is disregarding the central role of the singularity at the center of the black hole. It is commonly assumed that quantum gravity will substitute the singularity for *something* non-singular. This would clearly make information loss go away, but clarifying what is this *something* is fundamental to say to where information went. Currently, as far as the author knows, no proposed theory of quantum gravity (e.g., string theory, loop quantum gravity, minimal lengths scenarios) is close to this feat. Given the wide array of proposed ways to unify gravity and quantum mechanics and the astonishingly depressing lack of experimental guidance, sticking to the proven conservative approach of the semi-classical theory (which keep providing us with unexpected phenomena, see, e.g., [28]) might still hold important (and, at least in principle, verifiable) insights regarding quantum mechanical effects in curved space-times environments.

A Appendix 1: Alternative way of obtaining *u(v)*

We present here an alternative argument to that of chapter 5, based on [11], for discovering how the future modes (5.9) can be expressed in terms of the past modes (5.8). On one hand, this alternative derivation has the advantage of being simpler, but on the other hand, it describes a less physical situation than the one we used in the main body of this work. Apart from this fact, we assume that all hypotheses described earlier still hold (like asymptotic flatness and the geometrical optics approximation).

Consider a thin shell with mass M collapsing and forming a black-hole. Due to Birkhoff theorem, space-time outside the shell has the Schwarzschild metric and inside it is flat. Therefore, the radial part of both metrics can be written as

$$\mathrm{d}s_>^2 = \left(1 - \frac{2M}{r}\right)\mathrm{d}t^2 - \left(1 - \frac{2M}{r}\right)^{-1}\mathrm{d}r^2 \qquad r > R(T), \tag{A.1}$$

$$\mathrm{d}s_<^2 = \mathrm{d}T^2 - \mathrm{d}r^2 \qquad r \leq R(T), \tag{A.2}$$

where t is the proper time of an asymptotic observer, T is the proper time of an observer inside the shell and $R(t) = R(T(t))$ describes the evolution of the surface of the shell.

We will find relations between an in-going $v = t + r^* = const$ ray and an out-going $u(v) = t - r^* = const$ ray by linking them using the interior light-cone coordinates $U = T - r$, $V = T + r$ and the continuity condition for the metric at $R(T)$. The aforementioned condition, at $R(T)$, leads to the equality

$$1 - \left(\frac{\mathrm{d}R}{\mathrm{d}T}\right)^2 = \left(\frac{R - 2M}{R}\right)\left(\frac{\mathrm{d}t}{\mathrm{d}T}\right)^2 - \left(\frac{R - 2M}{R}\right)^{-1}\left(\frac{\mathrm{d}R}{\mathrm{d}T}\right)^2. \tag{A.3}$$

At the beginning of the collapse, far from $R = 2M$, $R(T)$ changes slowly and the above equation reduces to

$$1 \approx \left(\frac{\mathrm{d}t}{\mathrm{d}T}\right)^2. \tag{A.4}$$

And it is easily seen that $t(T)$ is an affine function.

Since we are far away from the surface, $r^* \approx r$, and we have

$$V = T + r \approx t + r^* + a = v + a, \tag{A.5}$$

where a is a constant. Now, at the center of the shell, $r = 0$ and we obtain immediately $U = V = v + a$. Finally, consider a $U = const$ ray that exit the shell as $R(T) \approx 2M$. We can expand $R(T)$ near T_0, where T_0 is the time of formation of the horizon (according to observers inside the shell), obtaining

$$R(T) = 2M + A(T_0 - T) + O(T^2). \tag{A.6}$$

where $A = const$. Substituting the above expansion in equation (A.3) and multiplying both sides by $\left(\frac{R-2M}{R}\right)$ leads to

$$\left(\frac{\mathrm{d}t}{dT}\right)^2 = \left(\frac{2M}{(T_0 - T)}\right)^2, \tag{A.7}$$

where we have considered only terms of order $(T - T_0)^{-1}$, which are the dominant ones. Integrating this gives us

$$t \approx -2M \log\left(\frac{T_0 - T}{b}\right), \tag{A.8}$$

where $b > 0$ is a constant and the sign is chosen such that the horizon forms at $t \to +\infty$. It is the linear Taylor expansion that is responsible for the correct result of this simple model, since it should be valid for all cases where a horizon forms. Also, substituting (A.6) in the definition of r^* (considering only the dominant term)

$$r^* \approx 2M \log\left(\frac{A(T_0 - T)}{2M}\right). \tag{A.9}$$

Therefore, $u = t - r^* = -4M \log \frac{(T_0 - T)}{c}$, where c is constant. Since

$$\begin{aligned} U &= T - r = T - 2M - A(T_0 - T) \\ &= T - T_0 + T_0 - 2M - A(T_0 - T) \\ &= T - T_0 + U_0 - A(T_0 - T), \end{aligned} \tag{A.10}$$

with $U_0 = T_0 - 2M$, we get, using all the preceding relations

$$v + a \approx V = U \approx U_0 - d(T_0 - T) \approx U_0 - f \exp\left(-\frac{u}{4M}\right), \tag{A.11}$$

where d and f are other constants. This implies, finally that

$$v = U_0 - a - f \exp\left(-\frac{u}{4M}\right) = v_0 - f \exp\left(-\frac{u}{4M}\right), \tag{A.12}$$

where we have identified the first sum of constant factors with v_0. Identifying f with K and inverting the relation above leads us to equation (5.25) as desired.

References

[1] C. M. Will; *The Confrontation between General Relativity and Experiment*, Living Rev. Relativity 17, (2014), 4. URL (cited on 04/26/2016): http://www.livingreviews.org/lrr-2014-4

[2] B. P. Abbott et al.; *Observation of Gravitational Waves from a Binary Black Hole Merger*, Phys. Rev. Lett. 116, 061102 (2016)

[3] N. D. Birrel & P. C. W. Davies; Cambridge University Press, England (1984), *Quantum Field Theory in Curved Spaces*

[4] R. M. Wald; University Of Chicago Press, USA (1984), *General Relativity*

[5] S. W. Haking and G. F. R. Ellis; Cambridge University Press, England (1975), *The Large Scale Structure of Space-time*

[6] R. Horodecki, P. Horodecki, M. Horodecki & K. Horodecki; *Quantum entanglement*, Rev. Mod. Phys. 81, 865 (2009)

[7] M. A. Nielsen & I. L. Chuang; Cambridge University Press, England (2011), *Quantum Computation and Quantum Information*

[8] M. Srednicki; *Entropy and area*, Phys. Rev. Lett. 71, 666 (1993)

[9] K. Ried; *Can entanglement explain black hole entropy?*, Master thesis, (2011), URL (cited on 06/03/2016): http://hdl.handle.net/11449/91804

[10] S. N. Solodukhin; *Entanglement Entropy of Black Holes*, Living Rev. Relativity 14, (2011), 8. URL (cited on 07/28/2016): http://www.livingreviews.org/lrr-2011-8

[11] L.H. Ford; Lectures given at the IX Jorge Andre Swieca Summer School, Campos do Jordão, SP, Brazil, February 1997, and at Soochow University, Taipei, Taiwan, June 1997, *Quantum Field Theory in Curved Space-times*

[12] L. E. Parker & D. J. Toms; Cambridge University Press, England (2009), *Quantum Field Theory in Curved Spacetime*

[13] J. L. Ball, I. Fuentes-Schuller & F. P. Schuller *Entanglement in an expanding spacetime* Phys.Lett. A 359, 6 (2006)

[14] J. D. Bekenstein; *Black holes and the second law*, Lettere al Nuovo Cimento (1971-1985) August 1972, V. 4, I. 15 (1972)

[15] R. M. Wald; *The thermodynamics of black holes*, Living Rev. Relativity 4, (2001), 6. URL (cited on 05/21/2016): http://www.livingreviews.org/lrr-2001-6

[16] J. M. Bardeen, B. Carter, S. W. Hawking; *The four laws of black hole mechanics*, Comm. Math. Phys. 31 (1973), no. 2, 161–170.

[17] M. Visser; *Thermality of Hawking Flux*, arXiv:1409.7754v3 [gr-qc] (2015)

[18] C. Barceló, S. Liberati & M. Visser; *Analogue gravity*, Living Rev. Relativity, 14, (2011), 3. URL (cited on 03/05/2016): http://www.livingreviews.org/lrr-2011-3

[19] G.W. Gibbons & S. W. Hawking; *Action integrals and partition functions in quantum gravity* Phys. Rev. D 15, 2752 (1977)

[20] S. W. Hawking; *Breakdown of predictability in gravitational collapse*, Phys. Rev. D 14, 2460 (1976)

[21] R. M. Wald; University Of Chicago Press, USA (1984), *Quantum Field Theory in Curved Spacetime and Black Hole Thermodynamics*

[22] E. Bianchi & M. Smerlak; *Entanglement entropy and negative energy in two dimensions* Physical Review D 90, 041904 (2014)

[23] S. Hossenfelder & L. Smolin; *Conservative solutions to the black hole information problem* Phys. Rev. D 81, 064009 (2010)

[24] S. W. Hawking; *Information Preservation and Weather Forecasting for Black Holes* arXiv:1401.5761 [hep-th] (2014)

[25] L. Susskind & J. Lindesay; World Scientific Publishing Company, Singapore (2004), *An Introduction To Black Holes, Information And The String Theory Revolution: The Holographic Universe*

[26] A. Almheiri, D. Marolf, J. Polchinski & J. Sully; *Black Holes: Complementarity or Firewalls?*, arXiv:1207.3123v4 [hep-th] (2012)

[27] A. Perez; *No Firewalls in Quantum Gravity: the Role of Discreteness of Quantum Geometry in resolving the Information Paradox* arXiv: 1410.7062v1 [gr-qc] (2014)

[28] W. C. C. Lima, G. E. A. Matsas & D. A. T. Vanzella; *Awaking the Vacuum in Relativistic Stars*, Phys. Rev. Lett. 105, 151102 (2010)