The black hole information problem: a critical review

Jean Douçot
Département de physique, École Normale Supérieure, Paris, France,
CPhT, École polytechnique, Palaiseau, France.

June 20, 2017

Abstract

I review in a pedagogical way the black hole information problem. I begin by discussing the main concepts of quantum field theory in curved spacetime and present a derivation of the Hawking effect. I explain how this semiclassical picture leads to the information problem. I then review and analyse in a critical way the main ideas that are proposed to solve the problem, emphasizing the assumptions and the inner logic of each proposal, as well as the radical differences of perspectives between them.

Contents

Introduction

In 1975, Stephen Hawking showed that, when taking quantum effects into account, a black hole radiates energy to infinity [1]. Moreover, this radiation is similar to the one emitted by a black body at the Hawking temperature T_H. This remarkable discovery had far-reaching implications. Before Hawking's result, the laws of black hole mechanics, derived in classical general relativity, [2] had suggested an analogy between black hole mechanics and thermodynamics. In this analogy, the same quantity T_H could be formally identified to a temperature. The Hawking effect strikingly showed that T_H really is the temperature of the black hole, and indiacted that the laws of black hole mechanics may indeed be thermodynamical laws.

Another important consequence of Hawking's result is that it showed that black holes evaporate. Indeed, the Hawking radiation emitted by a black hole carries energy to infinity. The black hole will thus loose mass, and a simple estimate indicates that it will evaporate in a finite time. As noticed by Hawking [3], this seems to have a puzzling consequence. A crucial feature in the analysis that derives the Hawking effect, done in the framework of quantum field theory in curved spacetime, is that the interior and the exterior of the black hole are entangled. More precisely, to find the state of the exterior only, i.e. of the Hawking radiation, we have to trace out the degrees of freedom corresponding to the interior. When saying that the radiation is thermal, one means that the reduced density matrix of the exterior describes a thermal state, in particular it is a mixed state. But then, when the black hole has evaporated away, it seems that there is somehow nothing left in the black hole for the radiation to be entangled with, so we are left with Hawking radiation only, described by a mixed state. Hawking therefore concluded that the evaporation of the black hole induces an evolution from pure states to mixed states. Such an evolution is not reversible, in that sense it implies that "information is lost".

Many people were not happy with this conclusion. Indeed, it apparently contradicts the postulates of quantum mechanics, according to which the evolution of the state vector is unitary, and in particular cannot bring a pure state to a mixed state, i.e. information cannot be lost. This is the so-called black hole information problem. It is perhaps the physical situation where the tension between general relativity and quantum mechanics manifests itself the most, it has attracted a considerable amount of attention since Hawking original papers and is currently still actively debated, although without much consensus about its resolution.

Hawking's prediction that black hole radiate is robust, and there is little doubt that it is indeed the case. However there are many uncertainties about the final stages of the evaporation, since the size of the black hole then reaches the Planck length, and expected quantum gravity effects, whatever they are, will become important, so that the semiclassical picture will break down. Various scenarios have been proposed to evade Hawking's conclusion that information is lost. Today, motivated by ideas from string theory and AdS/CFT, the position that has become dominant is that unitarity has to be saved, and the "information" inside the black hole must be carried out by the Hawking radiation by some mechanism. However, other researchers believe in very different scenarios, and even among the proponents of unitary evaporation, there is no consensus on the mechanism that should restore the purity of the Hawking radiation. It is fair to say that the current status of the information problem is quite a big mess: dozens of different possibilities have been suggested, and researchers in the field often have difficulties to understand and to convince each other. In the last years, the firewall controversy has triggered a strong revival of the subject and has given rise to several hundreds of contradictory papers.

This work is the final report of a MSc research project supervised by Marios Petropoulos. Both of us are non-experts in the field, so the idea was to try to understand the lengthy debates about information paradox and to clarify the assumptions of the different lines of thought in order to get some clear ideas. In this text, I review in a pedagogical way the black hole information problem. I begin by discussing the main ideas of quantum field theory in curved spacetime and how they lead to the Hawking effect. I then review and analyse the main possibilities for the evolution of black holes. The different proposals strongly disagree on some basic assumptions, which manifestly often creates a lot of confusion among people. For this reason, I take care to clearly state the assumptions of the arguments and try to emphasize the inner logic of each proposal, before questioning it.

In this spirit, I think it is honest to tell from the beginning the impression that I got from this study, since it inevitably orients the way I am presenting things. What makes most sense to me is to think that, *within the framework of quantum field theory in curved spacetime* in which the Hawking radiation is derived, information is indeed lost because there is a singularity. But we shouldn't worry too much about that. Indeed, QFT in curved spacetime is only an approximate theory, where spacetime is treated as classical. The natural way for unitarity to be restored is in the framework of some "full theory" of quantum gravity, where the gravitational degrees of freedom would be quantized. For example, "quantum gravity effects" may resolve the singularity and lead to a global spacetime structure different from the semiclassical geometry obtained from QFT in curved spacetime. I am surprised that the dominant attitude is to want to find some way to restore unitarity *while keeping the semiclassical geometry* with an event horizon and a singularity, which we can expect to be modified by quantum gravity, by introducing some mechanism that violates QFT in curved spacetimes at places we expect it to be valid.

For other introductory reviews on the black hole information paradox, I recommend [4]. See also [5–7] for recent reviews from the unitary evaporation perspective. For further reading on quantum field theory in curved spacetime and the Hawking effect, I recommend the review [8] and the books [9, 10].

Outline

I have tried to make this work pedagogical and understandable for a reader who is familiar with general relativity and basic quantum field theory. While the arguments in the literature are often quite vague, I have done my best to write as clearly as possible. In section 1, I review the formalism of quantum field theory in curved spacetime, emphasizing the conceptual differences between flat and curved spacetime. In section 2, I briefly discuss classical black hole thermodynamics and describe the Hawking effect. I discuss in section 3 the question of black hole evaporation and information loss. In section 4, I review the main ideas of the line of thought that wants to keep unitarity, including complementarity and firewalls. I analyse the assumptions of the arguments and discuss their validity. Finally, in section 5, I discuss other scenarios which are less radical in the sense that they don't require the brakdown of semiclassical theory in low-curvature regions.

1 Quantum field theory in curved spacetime

A first step to combine ingredients from both general relativity and quantum mechanics is to study the propagation of quantum fields on a fixed curved background spacetime. This is the framework in which the Hawking effect is obtained. As we shall see, the key point is that when we quantize a field in curved spacetime, there is no natural observer-independent definition of particles : the same state of the quantum field can be the vacuum state for some observers and a particle state for other observers. This is what happens for the Hawking effect : the quantum field is in the vacuum state for observers at past infinity, whereas observers at future infinity see particles. I have relied heavily on [8] for writing this part.

1.1 Harmonic oscillator with time-dependent frequency

1.1.1 General framework

I will begin by discussing the quantum theory of a forced harmonic oscillator in a time-dependent potential. Indeed, since free fields are analogous to a collection of harmonic oscillators, this will allow us to understand in a more elementary context many essential features that appear in quantum field theory in curved spacetime.

We consider a one-dimensional harmonic oscillator of mass m in a time-dependent potential $V(t,x) = 1/2m\omega^2(t)x^2$ determined by the time-dependent frequency $\omega(t)$. In the classical theory, the dynamics is governed by the action

$$S = \int L\,dt, \qquad L = \frac{1}{2}m\dot{x}^2 - \frac{1}{2}m\omega^2(t)x^2. \tag{1}$$

The canonical momentum is $p = \partial L/\partial \dot{x} = m\dot{x}$, and the classical equation of motion is

$$\ddot{x} + \omega^2(t)x = 0. \tag{2}$$

This is a linear differential equation of order 2, so the space of real solutions is of dimension 2. If f_1 and f_2 form a basis of solutions, and $f = f_1 + if_2$, a general real solution can be written as

$$x(t) = f(t)a + \bar{f}(t)\bar{a}, \tag{3}$$

where a is a complex constant of integration.

The space of complex solutions admits a symplectic structure, given by the inner product

$$\langle f, g\rangle = i\,m\left(\bar{f}\partial_t g - \partial_t \bar{f} g\right). \tag{4}$$

This is well-defined, because if f and g are solutions of the equation of motion, the bracket $\langle f, g\rangle$ doesn't depend on the time t at which the right hand side is evaluated. One can always rescale the solution f to have the normalization $\langle f, f\rangle = 1$.

Now let us go to the quantum theory. In canonical quantization, we have to replace the canonical coordinates x, p by operators $\hat{x}, \hat{p}$ that satisfy the canonical commutation relation

$$[\hat{x}, \hat{p}] = i. \tag{5}$$

Here, the conjugate momentum is $p = m\dot{x}$, so (5) becomes $[\hat{x}, \partial_t \hat{x}] = i/m$. The standard way to achieve that is to replace equation (3) (with f normalized) by

$$\hat{x}(t) = f(t)\hat{a} + \bar{f}(t)\hat{a}^\dagger, \tag{6}$$

where the operator $\hat{a}$ and his hermitian conjugate $\hat{a}^\dagger$ satisfy the commutation relation

$$[\hat{a}, \hat{a}^\dagger] = 1. \tag{7}$$

It is important to see that the creation and annihilation operators $\hat{a}^\dagger$ and $\hat{a}$ depend on the choice of the solution f. They can be expressed from $\hat{x}$ and f by

$$a = \langle f, \hat{x}\rangle, \qquad \hat{a}^\dagger = \langle \bar{f}, \hat{x}\rangle. \tag{8}$$

Using this, we can build the Hilbert space of the quantum theory. We define a state $|0\rangle$ by demanding it to be normalized and to satisfy $a|0\rangle = 0$. For each positive integer n we define the state $|n\rangle$ by $|n\rangle = (1/\sqrt{n!})(\hat{a}^\dagger)^n|0\rangle$. It is an eigenstate of the "number operator" $\hat{N} = \hat{a}\hat{a}^\dagger$ with eigenvalue n. We define the Hilbert space of the theory to be the span of all those states.

A special case is when the oscillator is time-independent, i.e. ω is constant. In this situation it turns out that there is naturally a preferred normalized solution of the equation of motion. It is the so-called *positive frequency solution*

$$f(t) = \frac{1}{\sqrt{2m\omega}}e^{-i\omega t}. \tag{9}$$

This solution has the property that the states $|n\rangle$ defined as above are eigenstates of the Hamiltonian operator, and the state $|0\rangle$ is the lowest energy state in the theory. Indeed, with this choice of f, the Hamiltonian has the expression

$$H = \omega(\hat{a}^\dagger\hat{a} + \tfrac{1}{2}) = \omega(N + \tfrac{1}{2}), \tag{10}$$

so the state $|n\rangle$ is an eigenstate, with energy $\omega(n + \frac{1}{2})$.

In the case where ω is constant, the state $|n\rangle$ can thus be physically interpreted as a state with "n excitations" of the harmonic oscillator. A crucial remark is that for a general function $\omega(t)$ there is no naturally preferred choice for the solution f. In particular, the states $|n\rangle$ will in general not be eigenstates of the Hamiltonian, and the state $|0\rangle$ will not be a lowest energy state.

1.1.2 Particle creation

Let us now discuss the effect that is analogous to particle creation in curved spacetime. It will be easier to understand it first for a single harmonic oscillator, and the results that we will obtain will be used for the derivation of the Hawking effect.

We consider the situation where the frequency $\omega(t)$ is constant in the asymptotic past and the asymptotic future:

$$\omega(t) \xrightarrow[t\to-\infty]{} \omega_{in}, \qquad \omega(t) \xrightarrow[t\to+\infty]{} \omega_{out} \tag{11}$$

As a consequence, there is a physical ground state $|0_{in}\rangle$ in the asymptotic past, and another ground state in $|0_{out}\rangle$ in the asymptotic future. The key point is that those two states are *different*. In particular, $|0_{out}\rangle$ will be an *excited* state with respect to the Fock space associated to $|0_{in}\rangle$.

We would like to find the precise relation between $|0_{in}\rangle$ and $|0_{out}\rangle$. The creation and annihilation operators for the *in* and *out* Fock spaces are determined by the solutions of the equations of motion

$$f_{in}(t) \xrightarrow[t\to-\infty]{} \frac{1}{\sqrt{2m\omega_{in}}} e^{-i\omega_{in}t}, \qquad f_{out}(t) \xrightarrow[t\to-\infty]{} \frac{1}{\sqrt{2m\omega_{out}}} e^{-i\omega_{out}t} \tag{12}$$

Since the space of solutions is two-dimensional, there exist two complex numbers α and β such that

$$f_{out} = \alpha f_{in} + \beta \bar{f}_{in}. \tag{13}$$

The normalization $\langle f_{out}, f_{out}\rangle = 1$ gives

$$|\alpha|^2 - |\beta|^2 = 1. \tag{14}$$

Using (8), we get the expression

$$a_{out} = \alpha a_{in} - \bar{\beta} a^\dagger_{in} \tag{15}$$

between the *in* and *out* creation and annihilation operators (from now on we drop the hats on the operators). This relation is called a *Bogoliubov transformation*, the numbers α and β are called *Bogoliubov coefficients*. From equation (15) we get the mean number of *out* excitations in the *in* vacuum:

$$\langle 0_{in}| a^\dagger_{out} a_{out} |0_{in}\rangle = |\beta|^2. \tag{16}$$

We want to express $|0_{in}\rangle$ in the basis of *out* eigenstates $|n_{out}\rangle$. Using again (8) we get another expression similar to (15):

$$a_{in} = \alpha a_{out} + \bar{\beta} a^\dagger_{out}. \tag{17}$$

Then, $a_{in}|0_{in}\rangle = 0$ implies

$$a_{out}|0_{in}\rangle = -\frac{\bar{\beta}}{\alpha} a^\dagger_{out}|0_{in}\rangle. \tag{18}$$

The solution of this equation is

$$|0_{in}\rangle = \mathcal{N} \exp\left[-\frac{\bar{\beta}}{2\alpha} a^\dagger_{out} a^\dagger_{out}\right] |0_{out}\rangle = \mathcal{N} \sum_n \frac{\sqrt{2n!}}{n!} \left(-\frac{\bar{\beta}}{2\alpha}\right)^n |2n_{out}\rangle. \tag{19}$$

This can be suggested by noticing that the commutation relation $[a_{out}, a^\dagger_{out}] = 1$ is analogous to $[\partial_x, x] = 1$, so we may imagine that $a_{out} = \partial/\partial a^\dagger_{out}$. We can then view the equation as a simple differential equation. The constant $\mathcal{N}$ is fixed by the normalization condition, it is found to be $\mathcal{N} = |\alpha|^{-1/2}$.

1.2 Quantum field theory

We now come to the formulation of quantum field theory in curved spacetime. To fix the ideas and see what changes when spacetime is curved, I will first briefly review what happens in flat spacetime.

1.2.1 Flat spacetime

We will restrict to the case of a free massive scalar field, which already contains all the essential ingredients. For higher spin fields and interacting fields, things are more technical but the basic ideas remain the same. We use the signature $(+---)$.

In the classical theory, the dynamics of a massive scalar field ϕ of mass m is governed by the action

$$S = \frac{1}{2}\int d^4x\,(\partial^2\phi - m^2\phi^2). \tag{20}$$

The classical equation of motion is the Klein-Gordon equation

$$(\partial_\mu\partial^\mu + m^2)\phi = 0. \tag{21}$$

In Fourier space, the Klein-Gordon equation becomes a linear combination of ordinary differential equations for each wave vector $\mathbf{k}$. A complete set of solutions is given by the positive frequency normalized modes

$$f_{\mathbf{k}} = \frac{1}{\sqrt{2\omega(\mathbf{k})}(2\pi)^{3/2}}e^{-i\omega(\mathbf{k})t}e^{i\mathbf{k}.\mathbf{x}}, \tag{22}$$

with $\omega(\mathbf{k}) = \sqrt{\mathbf{k}^2 + m^2}$, together with their complex conjugates $\bar{f}_{\mathbf{k}}$. A general real solution can be expressed as

$$\phi = \frac{1}{(2\pi)^{3/2}}\int d^3\mathbf{k}\left(f_{\mathbf{k}}a_{\mathbf{k}} + \bar{f}_{\mathbf{k}}\bar{a}_{\mathbf{k}}\right). \tag{23}$$

where the $a_{\mathbf{k}}$ are complex integration constants.

To canonically quantize the theory, we promote the numbers $a_{\mathbf{k}}$ and $\bar{a}_{\mathbf{k}}$ to operators $\hat{a}_{\mathbf{k}}$ and $\hat{a}^\dagger_{\mathbf{k}}$ that satisfy the commutation relations $[\hat{a}^\dagger_{\mathbf{k}}, \hat{a}_{\mathbf{l}}] = \delta_{\mathbf{k},\mathbf{l}}$. The field operator $\hat{\phi}$ defined by the mode expansion

$$\hat{\phi} = \frac{1}{(2\pi)^{3/2}}\int d^3\mathbf{k}\left(f_{\mathbf{k}}\hat{a}_{\mathbf{k}} + \bar{f}_{\mathbf{k}}\hat{a}^\dagger_{\mathbf{k}}\right) \tag{24}$$

then satisfies the canonical commutation relation $[\hat{\phi}(t,\mathbf{x}), \hat{\pi}(t,\mathbf{y})] = i\delta(\mathbf{x}-\mathbf{y})$.

The Hilbert space of the theory is constructed as follows. The vacuum state $|0\rangle$ is defined by the property that it is annihilated by all $\hat{a}_{\mathbf{k}}$. We then construct other states $\hat{a}^\dagger_{\mathbf{k_1}}\ldots\hat{a}^\dagger_{\mathbf{k_n}}|0\rangle$ by acting with creation operators, they are interpreted as states with n particles. The total Hilbert space is the span of all those states.

This formulation of quantum field theory is the usual one, but it relies heavily on the simple form of the Minkowski metric which allows us to split all quantities into temporal and spatial parts, so it is not adapted for generalizing to curved spacetime. Let us discuss the reformulation that will carry over to curved spacetime.

The space of solutions of the classical Klein-Gordon equation has a symplectic structure, with the inner product defined by

$$\langle f, g\rangle = i\int_{\Sigma_t}\bar{f}\partial_t g - (\partial_t\bar{f})g \tag{25}$$

on any spatial slice Σ_t of constant time. We notice that the inner products of the mode function $f_{\mathbf{k}}$ and their conjugates are

$$\langle f_{\mathbf{k}}, f_{\mathbf{l}}\rangle = \delta_{\mathbf{k},\mathbf{l}}, \tag{26}$$

$$\langle \bar{f}_{\mathbf{k}}, \bar{f}_{\mathbf{l}}\rangle = -\delta_{\mathbf{k},\mathbf{l}}, \tag{27}$$

$$\langle f_{\mathbf{k}}, \bar{f}_{\mathbf{l}}\rangle = 0. \tag{28}$$

Let us denote by $\mathcal{S}_p$ the span of the positive frequency modes $f_{\mathbf{k}}$. The above relations mean that the space $\mathcal{S}$ of solutions of the Klein-Gordon equation decomposes with respect to the inner product as the orthogonal direct sum

$$\mathcal{S} = \mathcal{S}_p \oplus \bar{\mathcal{S}}_p, \tag{29}$$

where the elements of $\mathcal{S}_p$ have positive norm. The annihilation operators can be expressed from $\hat{\phi}$ and the $f_{\mathbf{k}}$ using the inner product as

$$\hat{a}_{\mathbf{k}} = \langle f_{\mathbf{k}}, \hat{\phi} \rangle. \tag{30}$$

The Hilbert space is then defined as before using the creation and annihilation operators associated to the $f_{\mathbf{k}}$.

1.2.2 Curved spacetime

We will now see how to define quantum field theory in curved spacetime. We consider a free massive scalar field in a spacetime with metric $g_{\mu\nu}$. The classical action is

$$S = \frac{1}{2} \int d^4x \sqrt{|g|} (g^{\mu\nu} \partial_\mu \partial_\nu \phi - m^2 \phi^2), \tag{31}$$

which gives the classical equation of motion

$$(\Box + m^2)\phi = 0, \tag{32}$$

where $\Box = |g|^{-1/2} \partial_\mu |g|^{1/2} |g^{\mu\nu} \partial_\nu$.

To quantize the theory, we need to foliate the spacetime into spatial slices. As a consequence, the quantum field theory can be defined only is the spacetime is *globally hyperbolic*, i.e. if it admits a global "time function" such that it is foliated by spacelike surfaces Σ_t of "constant time". Let us assume that it is the case and choose adapted coordinates $x^\mu = (x^0, x^i)$, such that x^0 is the time coordinate and x^i are the spatial coordinates. We can rewrite the action as

$$S = \int dx^0 L, \tag{33}$$

where the Lagrangian is $L = \int dx^3 \mathcal{L}$, with $\mathcal{L}$ denoting the Lagrangian density. The canonical momentum at time x^0 is then defined by

$$\pi(x^i) = \frac{\delta L}{\delta(\partial_0 \phi)} = |g|^{1/2} g^{\mu 0} \partial_\mu \phi(x^0, x^i). \tag{34}$$

To quantize, we have to replace ϕ and π by operators $\hat{\phi}$ and $\hat{\pi}$ that satisfy the canonical commutation relation at equal times

$$[\hat{\phi}(x^i), \hat{\pi}(y^j)] = i\delta(x^i - y^i). \tag{35}$$

We could try to do a mode decomposition, but in a generic curved spacetime there is no natural physically preferred choice. The reformulation presented above will allow us to formulate the quantum field theory in a nice way, without using coordinates.

As in the flat case, the space of classical solutions possesses a symplectic structure given by the inner product

$$\langle f, g \rangle = i \int_\Sigma d\Sigma_\mu |g|^{1/2} g^{\mu\nu} \left(\bar{f} \partial_\nu g - (\partial_\nu \bar{f}) g \right), \tag{36}$$

where Σ is any Cauchy surface (the inner product is independent on Σ). Let $\hat{\phi}$ a hermitian field operator, i.e. an operator-valued distribution on spactime. For each solution f of the Klein-Gordon equation, we can use the analogue of equations (8) and (30) to construct an annihilation and a creation operator associated to f and $\hat{\phi}$:

$$\hat{a}(f) = \langle f, \hat{\phi} \rangle, \qquad \hat{a}^\dagger(f) = -\hat{a}(\bar{f}). \tag{37}$$

The field operator $\hat{\phi}$ satisfies the canonical commutation relation (35) if and only if the creation and annihilation operators satisfy the commutation relation

$$[\hat{a}(f), \hat{a}^\dagger(g)] = \langle f, g \rangle. \tag{38}$$

Notice that this expression does neither involve coordinates nor an explicit foliation of spacetime. It will therefore provide the good way to quantize the theory.

To formulate the quantum field theory we need to have a decomposition of the space $\mathcal{S}$ of solutions of the Klein-Gordon equation as an orthogonal direct sum

$$\mathcal{S} = \mathcal{S}_p \oplus \bar{\mathcal{S}}_p, \tag{39}$$

such that all elements of $\mathcal{S}_p$ have positive norm: $\langle f, f \rangle > 0$ for all $f \in \mathcal{S}_p$. In the flat case, the natural choice for S_p was the space of positive frequency solutions. In curved spacetime, there is in general no such natural choice. It is possible to show under general hypotheses that such a "positive frequency subspace" S_p always exists, but it is not unique [9]. Let us assume we have chosen some $\mathcal{S}_p$. To each f in $\mathcal{S}_p$ we associate an annihilation operator $\hat{a}(f)$, such that the $\hat{a}(f)$ satisfy the algebra (38). Taking complex conjugates of this relation, we also get that for any f, g in $\mathcal{S}_p$, we have $[\hat{a}(f), \hat{a}(g)] = [\hat{a}^\dagger(f), \hat{a}^\dagger(g)] = 0$. We then construct the Hilbert space of the theory as follows. We define the state $|0\rangle$ to be annihilated by $\hat{a}(f)$ for all f in $\mathcal{S}_p$. The Hilbert space is the span of all states of the form

$$\hat{a}^\dagger(f_1) \dots \hat{a}^\dagger(f_n)|0\rangle, \tag{40}$$

where $f_1, \dots, f_n$ are in $\mathcal{S}_p$. The field operator $\hat{\phi}$ such that $\hat{a}(f) = \langle f, \hat{\phi} \rangle$ for all f in $\mathcal{S}_p$ then satisfies the canonical commutation relation. Notice that the Hilbert space that we obtain, i.e. the representation of the canonical commutation relations, depends on the choice of $\mathcal{S}_p$. In general the representations obtained from different $\mathcal{S}_p$ will be unitarily inequivalent. Different reprentations are related by a (formal) Bogoliubov transformation.

2 The Hawking effect

We now turn to the analysis of the Hawking effect. Before presenting a derivation of Hawking's result, I will recall a few important facts about black holes.

2.1 Black holes

2.1.1 The Schwarzschild geometry

We will only consider a non-rotating Schwarzschild black hole, for which the metric is given in spherical coordinates by

$$ds^2 = \left(1 - \frac{2m}{r}\right) dt^2 - \left(1 - \frac{2m}{r}\right)^{-1} dr^2 - r^2(d\theta^2 + \sin^2\theta d\phi^2). \tag{41}$$

The Schwarzschild metric presents a singularity at $r = 0$. The surface $r = 2m$ is called the horizon of the black hole, it is the boundary of the black hole region $r < 2m$. It is an *event horizon*, because a timelike or null geodesic that enters the black hole region cannot escape to future infinity and ends on the singularity. It is also a *Killing horizon*, because the stationary Killing vector field ∂_t is null on the horizon. The black hole horizon is therefore a null surface, it is also a surface of infinite redshift with respect to observers far away from the black hole.

We will use other coordinates associated with the behaviour of radial null geodesics. The "tortoise coordinate" is defined $r_* = r + 2m \ln(r/2m - 1)$. The retarded and advanced time coordinates are defined by

$$u = t - r_*, \tag{42}$$
$$v = t + r_*. \tag{43}$$

Outgoing null geodesics are given by $u = const$, and ingoing null geodesics by $v = const$.

The most convenient way to represent the black hole geometry is to use a Penrose diagram, obtained by a conformal transformation. When dealing with black holes, one very often considers the *extended Schwarzschild spacetime* which is obtained by extending all geodesics in such a way that each geodesic either goes to infinity or ends on a singularity both in the past and in the future. It is larger than

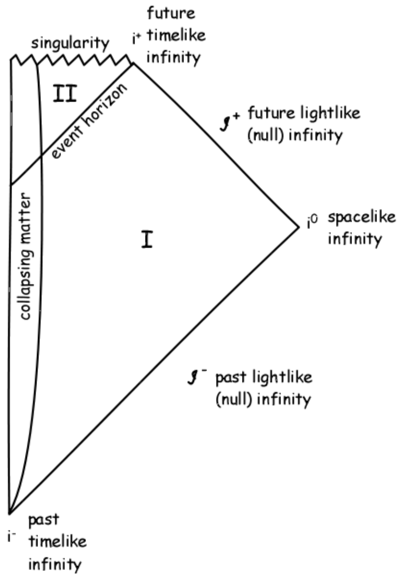


Figure 1: Penrose diagram of a black hole formed from the collapse of a massive object.

the patch covered by the spherical coordinates (t, r, θ, ϕ), and presents two new regions : a second asymptotically flat region and a white hole region.

However, the extended Schwarzschild geometry does not describe real astrophysical black holes that form from the collapse of massive object. In the following, we will always consider the geometry of a black hole created from collapse, which is represented by the "one-sided" Penrose diagram shown on Fig. 1. The geometry outside the matter cloud is given by the Schwarzschild metric.

2.1.2 Classical black hole thermodynamics

Let us say a few words about the classical laws of black hole mechanics. Historically, they were found before the Hawking effect was known and already suggested an analogy between black holes and thermodynamics. The laws of black hole mechanics are derived in classical general relativity for general black holes. In a generic asymptotically flat spacetime, a black hole is defined as a region from which light cannot escape to infinity. More precisely, the black hole region is the region of spacetime that is not contained in the causal past of future infinity $I^-(\mathscr{I}^+)$. The *event horizon* is defined as the boundary of the black hole region, it is thus a global property of the spacetime.

An important property of the event horizon of a black hole is that it is also a *Killing horizon*. This means that the horizon it is a null hypersurface, and there exists a Killing field ξ to which it is orthogonal. For any Killing horizon, by definition we have $\xi^\mu \xi_\mu = 0$ at the horizon. As a consequence, the vector $\nabla^\mu(\xi^\alpha \xi_\alpha)$ is orthogonal to the horizon. Since ξ^μ is also orthogonal to the horizon, those two vectors must be proportional, there exists a function κ defined on the horizon, called *surface gravity*, such that

$$\nabla^\mu(\xi^\alpha \xi_\alpha) = -2\kappa \xi^\mu. \tag{44}$$

For a black hole, the surface gravity is constant on the horizon. This property is called the *zeroth law* of black hole mechanics, in analogy with the zeroth law of thermodynamics that states that a system in equilibrium has constant temperature.

The *area law* states that the area of the event horizon of a black hole always increases under any classical process. This property is formally similar to the second law of thermodynamics, according to which the entropy of a closed system always increases. This suggests that we could associate to a black hole an entropy proportional to its area.

Remarkably, a stationary black hole is characterized by only three parameters : its mass M, angular momentum J and electric charge Q. This is the so-called no-hair theorem. Two nearby

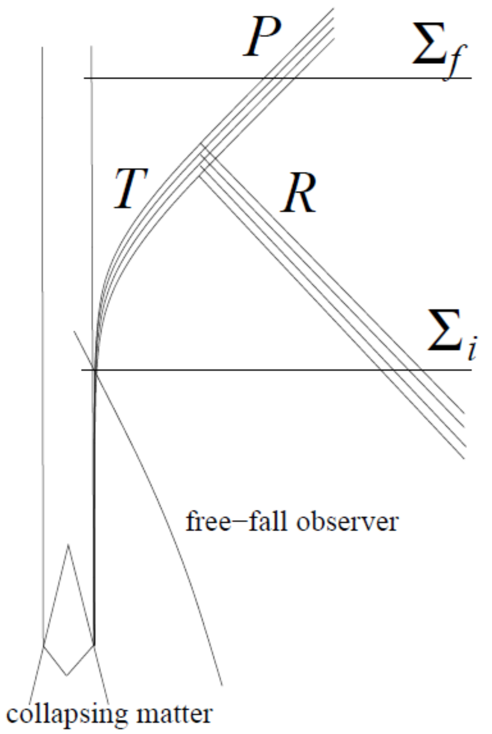


Figure 2: A spacetime diagram of a black hole formed from collapse, represented in ingoing Eddington-Finkelstein coordinates. P is an outgoing positive Killing frequency wavepacket, it splits into a transmitted parts T and a reflected part R when we propagate it backwards from the slice Σ_f to the slice Σ_i (figure taken from [8]).

stationary black hole configurations are related by

$$dM = \frac{\kappa}{8\pi} dA + \Omega dJ + \Phi dQ, \tag{45}$$

where Ω and Φ are the angular velocity and the electric potential of the horizon. This relation is called the *first law* of black hole mechanics, in analogy with the first law of thermodynamics. From those laws we see that there is a formal analogy between black hole mechanics and thermodynamics, in which the surface gravity corresponds to the temperature, and the area corresponds to the entropy.

2.2 Hawking radiation

Let us consider a quantum field propagating in the geometry of a black hole formed from collapse. In the asymptotic past and the asymptotic future of the black hole, spacetime is flat and there is a natural definition of particles. However, those definitions have no reason to coincide, and indeed they don't. In particular, a state of the quantum field with no ingoing particles is seen by an observer at future infinity as a particle state. Hawking showed that remarkably, for observers at future infinity, this radiation emitted by the black hole is a thermal radiation at the temperature $T_H = \kappa/2\pi$. More precisely, the state of the quantum field is such that the interior and the exterior of the black hole are *entangled*, and its restriction to the exterior is given by a thermal density matrix. In this section I will sketch the derivation of the Hawking effect and discuss the entangled nature of the quantum state.

2.2.1 Number of outgoing particles

We consider a free massive scalar field propagating in the spacetime of a black hole formed from collapse. Let $|\Psi\rangle$ the state of the quantum field. We would like to compute the number of outgoing particles seen by an observer at future infinity. To formulate this, we consider a wavepacket P, i.e. a solution of the classical Klein-Gordon equation, that is outgoing with positive Killing frequency far away from the black hole (by Killing frequency we mean the frequency with respect to the Killing time t). We want to compute the average number of corresponding particles $\langle\Psi|a^\dagger(P)a(P)|\Psi\rangle$ (we drop the hat on the operators). We will use two Cauchy surfaces, an early slice Σ_i and a late slice Σ_f, to "catch" the wavepacket P, as illustrated on Fig 2. The derivation will rely in a crucial way on the hypothesis, motivated by the equivalence principle, that an observer falling into the black hole will see no excitations at the horizon. For this reason we consider a freely-falling observer with proper time τ, who crosses the horizon at $\tau = 0$ on Σ_i (see Fig 2).

On the late time slice Σ_f the wavepacket P is far away from the black hole. When we propagate P back to the early slice Σ_i it decomposes into a "reflected" part R and a "transmitted" part T that approaches the horizon. We have

$$P = R + T. \tag{46}$$

The annihilation operator associated to P is given by $a(P) = \langle P, \phi\rangle$, where ϕ is the field operator corresponding to $|\Psi\rangle$. In this expression the inner product can be evaluated on Σ_i, which implies

$$a(P) = a(R) + a(T). \tag{47}$$

We assume that the state $|\Psi\rangle$ contains no incoming excitations. Since the metric is stationary, the reflected part R has the same positive Killing frequency as T. This implies that R has positive frequency in the asymptotic past. Therefore, the state $|\Psi\rangle$ is annihilated by $a(R)$. Then equation (47) implies that

$$\langle\Psi|a^\dagger(P)a(P)|\Psi\rangle = \langle\Psi|a^\dagger(T)a(T)|\Psi\rangle. \tag{48}$$

Now, the wavepacket T has not positive frequency with respect to the proper time τ of the infalling observer. T is decomposed into a positive and a negative frequency part,

$$T = T^+ + T^-. \tag{49}$$

This implies

$$a(T) = a(T^+) + a(T^-) = a(T^+) - a^\dagger(\overline{T^-}). \tag{50}$$

The mode T is blueshifted at the horizon when propagated backwards, therefore T^+ and T^- are very high frequency with respect to τ.

At this point, we use the hypothesis that the infalling observer sees no high-frequency excitations, i.e. the modes with frequency higher than $1/2m$ should be unexcited. This means that

$$a(T^+)|\Psi\rangle = 0, \qquad a(\overline{T^-})|\Psi\rangle = 0. \tag{51}$$

Taking this into account and using (50), we get that

$$\langle\Psi|a^\dagger(P)a(P)|\Psi\rangle = \langle\Psi|[a^(\overline{T^-}), a^\dagger(\overline{T^-})]|\Psi\rangle = \langle\overline{T^-}, \overline{T^-}\rangle = -\langle T^-, T^-\rangle. \tag{52}$$

We have thus reduced the problem to computing the norm of the negative frequency part of T with respect to τ.

To find T^-, we must determine the behaviour of the wavepacket T close to the horizon. The scalar field ϕ satisfies the Klein-Gordon equation $(\Box + m^2)\phi = 0$, it can be decomposed in spherical harmonics

$$\phi(t, r, \theta, \phi) = \sum_{lm} \frac{\phi_{lm}}{r} Y_{lm}(\theta, \phi), \tag{53}$$

where the mode $\phi_{lm}(t, r)$ satisfies

$$(\partial_t^2 - \partial_{r_*}^2 + V_{lm})\phi_{lm} = 0, \tag{54}$$

with V_{lm} an effective potential given by

$$V_{lm}(r) = \left(1 - \frac{2m}{r}\right)\left(\frac{2m}{r^3} + \frac{l(l+1)}{r^2} + m^2\right). \tag{55}$$

Close to the horizon, for $r \to 2m$, one has $r_* \sim \ln(\frac{r}{2m} - 1) \to -\infty$, therefore $V_{lm}(r_*) \sim \exp(r_*/2m)$ is negligible in the coordinates (t, r_*). This means that $\phi_{lm}(t, r_*)$ satisfies the massless wave equation close to the horizon. The general solution is thus of the form $\phi(t, r_*) = f(u) + g(v)$. Now, the wavepacket P is outgoing and has positive frequency far away from the black hole, this implies that the wavepacket T can only depend on the retarded time $u = t - r_*$. Assuming that far away P is concentrated around some Killing frequency ω, then $P_{lm} \sim \exp(-i\omega t)$. Since the metric is stationary the t-dependence remains the same when we propagate backwards, so we must have $T_{lm} \sim \exp(-i\omega u)$ close to the horizon. We must relate this to the proper time τ associated to the freely falling geodesic. Close to the horizon we have $\tau \sim \exp(-\kappa u)$, with $\kappa = 1/4m$ the surface gravity of the Schwarzschild black hole. We thus have

$$T \sim \begin{cases} \exp\left(i\frac{\omega}{\kappa}\ln(-\tau)\right) & \text{if } \tau < 0, \\ 0 & \text{if } \tau > 0. \end{cases} \tag{56}$$

We have to find the positive and negative frequency part of T. One can show [8] that the Bogoliubov transformation is given by

$$T^+ = c^+(T + e^{-\pi\omega/\kappa\tilde{T}}), \tag{57}$$

$$T^- = c^-(T + e^{+\pi\omega/\kappa\tilde{T}}), \tag{58}$$

where the wavepacket $\tilde{T}$ is the "mirror mode" of T inside the horizon, i.e. it is constant on outgoing null rays, and $\tilde{T}(t) = T(-\tau)$ for $\tau > 0$. The constants are found to be $c^- = (1 - e^{2\pi\omega/\kappa})^{-1}$ and $c^+ = -e^{2\pi\omega/\kappa}c^-$. Using that $\langle T, \tilde{T}\rangle = 0$ since T and $\tilde{T}$ don't overlap and that $\langle \tilde{T}, \tilde{T}\rangle = -\langle T, T\rangle$ because T and $\tilde{T}$ have opposite τ-dependences , we finally get that the mean number of observed particles is

$$\langle\Psi|a^\dagger(P)a(P)|\Psi\rangle = -\langle T^-, T^-\rangle = \frac{\langle T, T\rangle}{e^{2\pi\omega/\kappa} - 1}. \tag{59}$$

The factor $\exp(2\pi\omega/\kappa) - 1$ is the mean number of excitations for a thermal state at Hawking's temperature $T_H = \hbar\kappa/2\pi$ (restoring Planck's constant). It is multiplied in (59) by the so-called *greybody factor* $\Gamma = \langle T, T\rangle$. Γ can be interpreted as the probability for the wavepacket P to be "transmitted" rather than "reflected" when we propagate it backwards. Eq. (59) makes the classical analogy with thermodynamics more precise: T_H really is the physical temperature of the black hole, it is proportional to the surface gravity of the black hole as expected from the classical analogy. Now, the second law of black hole mechanics gives the proportionality factor for the entropy:

$$S_{BH} = \frac{A}{4\hbar}. \tag{60}$$

This is the *Bekenstein-Hawking entropy* of the black hole. Under some hypotheses one can show [9] that a *generalized second law* holds: the total entropy $S_{matter} + S_{BH}$ never decreases.

2.2.2 The quantum state and its entangled nature

We would now like to find the expression of the quantum state, which will give us much more information that the expectation value we computed in the last paragraph. We look at partner modes close to the horizon such as T and $\tilde{T}$ in the previous analysis. Since the field is non-interacting, the full quantum state will be the tensor product of the states of each pair of modes.

The Bogoliubov transformation (58) allows us to write the "local vacuum" conditions (51) as

$$\left[a(T) - e^{-\pi\omega/\kappa}a^\dagger(\bar{\tilde{T}})\right]|\Psi\rangle = 0, \tag{61}$$

$$\left[-a^\dagger(T) + e^{+\pi\omega/\kappa}a(\bar{\tilde{T}})\right]|\Psi\rangle = 0. \tag{62}$$

Notice that the modes T and $\bar{\tilde{T}}$ are positive frequency with respect to τ. Let us introduce the state $|B\rangle$ in the Fock space of T and $\bar{\tilde{T}}$ that is annihilated by $a(T)$ and $a(\bar{\tilde{T}})$, i.e.

$$a(T)|B\rangle = 0 \tag{63}$$

$$a(\bar{\tilde{T}})|B\rangle = 0. \tag{64}$$

This state is called the *Boulware vacuum* associated to those modes. Then, in analogy with the expression (19) that we had for the time-dependent harmonic oscillator, we get the expression of the part $|U\rangle$ of the quantum state associated to those modes, called the *Unruh vacuum*:

$$|U\rangle \propto \exp\left[e^{-\pi\omega/\kappa} a^\dagger(\hat{T}) a^\dagger(\hat{\bar{\bar{T}}})\right] |B\rangle, \tag{65}$$

where $\hat{T}$ and $\hat{\bar{\bar{T}}}$ are the normalized wavepackets associated to T and $\bar{\bar{T}}$.

We will now discuss the structure of this quantum state. In particular, we will see that it is such that the interior and the exterior of the black hole are *entangled*. This is the crucial point that leads to the information problem.

I will first recall a few facts about entanglement. In general quantum mechanics, entanglement occurs in a *bipartite system*, i.e. a system whose Hilbert space factorizes as a tensor product

$$\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B. \tag{66}$$

The states in $\mathcal{H}$ of the form $|\psi\rangle = |\psi_A\rangle \otimes |\psi_B\rangle$ are called *product states*. There are states in $\mathcal{H}$ that cannot be decomposed as a product state, they are said to be *entangled*. If $|\psi\rangle$ be a state in $\mathcal{H}$, we define its restriction on the subsystem A (respectively B) by taking the partial trace with respect to B (respectively A):

$$\rho_A = \mathrm{Tr}_B(|\psi\rangle\langle\psi|), \qquad \rho_B = \mathrm{Tr}_A(|\psi\rangle\langle\psi|). \tag{67}$$

If $|\psi\rangle$ is product state, we have $\rho_A = |\psi_A\rangle\langle\psi_A|$, i.e. the reduced density matrix ρ_A is a pure state. On the other hand when $|\psi\rangle$ is entangled, ρ_A is a mixed state.

The *von Neumann entropy* measures how much a quantum state is mixed. The von Neumann entropy of a quantum state in a system $\mathcal{H}$ given by a density matrix ρ is defined as

$$S_{vN}(\rho) = -\,\mathrm{Tr}(\rho \ln \rho). \tag{68}$$

If ρ is a pure state (i.e. if it is of the form $\rho = |\psi\rangle\langle\psi|$) then $S_{vN}(\rho) = 0$, if ρ is mixed then $S_{vN}(\rho) > 0$. The von Neumann entropy is maximal when $\rho = \frac{1}{N} I$ with $N = \dim \mathcal{H}$, in this case $S_{vN}(\rho) = \ln N$. Let $|\psi\rangle$ be a state of a bipartite system $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$. Its entanglement entropy is defined as

$$S_{ent}(\rho) = S_{vN}(\rho_A) = S_{vN}(\rho_B). \tag{69}$$

If ρ is a product state, then $S_{ent}(\rho) = 0$, and if ρ is entangled, then $S_{ent}(\rho) > 0$. The entanglement entropy is a measure of how much the subsystems A and B are entangled in the quantum state $|\psi\rangle$.

Let us now see how to formulate entanglement in the framework of quantum field theory in curved spacetime. Recall that the Hilbert space of the theory is constructed as the Fock space built from "positive frequency" solutions of the classical Klein-Gordon equation. Let Σ be a Cauchy surface in our spacetime, let us divide it in two disjoint parts Σ_1 and Σ_2, so that $\Sigma = \Sigma_1 \cup \Sigma_2$. This decomposition of Σ induces a decomposition of the space of classical solutions $\mathcal{S}$ into the space $\mathcal{S}_1$ of solutions with support in Σ_1 and the space $\mathcal{S}_2$ of solutions with support in Σ_2. We have $\mathcal{S} = \mathcal{S}_1 \oplus \mathcal{S}_2$, and $\mathcal{S}_1$ and $\mathcal{S}_2$ are orthogonal with respect to the inner product. This gives a tensor product decomposition of the Hilbert space: $\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2$, where $\mathcal{H}_1$ and $\mathcal{H}_2$ are the Fock spaces built from "positive frequency" solutions in $\mathcal{S}_1$ and $\mathcal{S}_2$. This allows us to define the restriction of the quantum state to Σ_1 and Σ_2 and the corresponding entanglement entropy.

With this in mind, let us go back to our analysis of the Hawking effect and look at the quantum state (65). The Cauchy surface Σ_i can be divided into the part inside the black hole Σ_{in} and the part outside Σ_{out}. The mode T has support on Σ_{out}, whereas $\bar{\bar{T}}$ has support on Σ_{in}. Let us denote $|n_{out}\rangle$ and $|n_{int}\rangle$ the states with n excitations of T and $\bar{\bar{T}}$ respectively. The Unruh vacuum $|U\rangle$ is then

$$|U\rangle \propto \sum_n e^{-n\pi\omega/\kappa} |n_{in}\rangle |n_{out}\rangle. \tag{70}$$

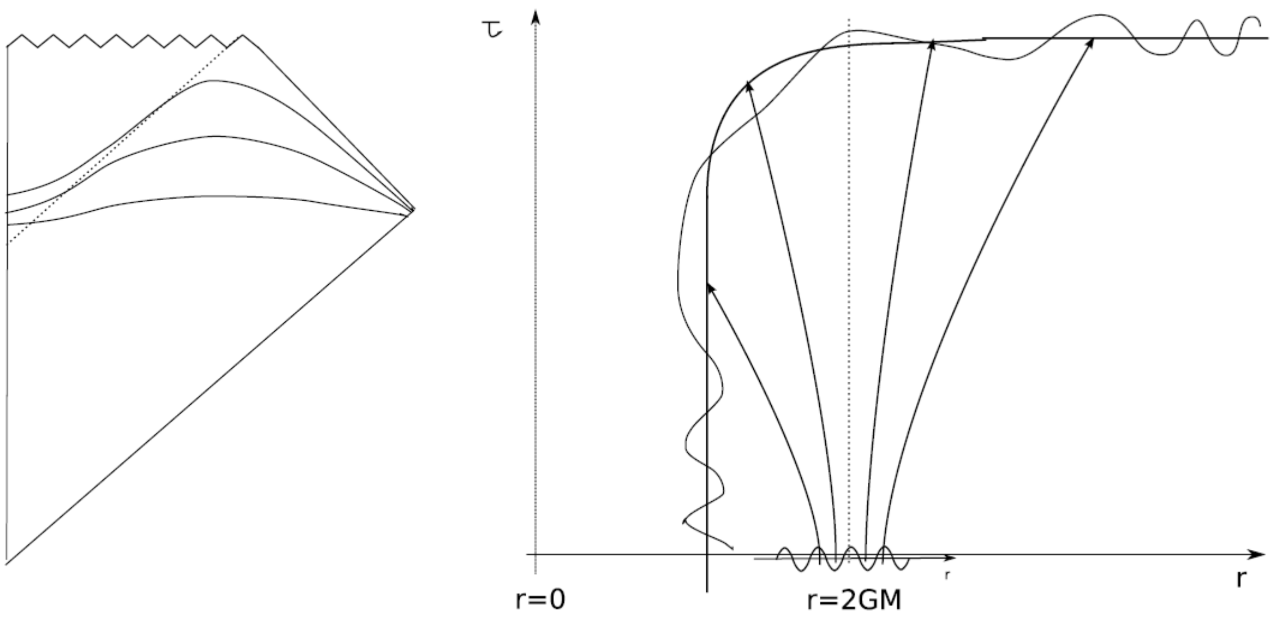
The reduced density matrix on Σ_{out}, which describes the state of the Hawking radiation, is thus

$$\rho_{out} = \mathrm{Tr}_{in}(|U\rangle\langle U|) \propto \sum_n e^{-2n\pi\omega/\kappa} |n_{out}\rangle\langle n_{out}|. \tag{71}$$

This state is a *mixed* state, more precisely, it is a thermal state at the Hawking temperature T_H. This means that the Hawking radiation, i.e the modes outside the horizon, are *entangled* with the modes inside the horizon. An important remark is that this entangled structure of the quantum state is a generic feature of quantum field theory. In particular, for flat Minkowski space, if we divide any spatial slice Σ into two pieces Σ_1 and Σ_2, these two pieces are always entangled when the field is in the vacuum state.

2.3 The nice slice picture

There is an interesting intuitive picture of the Hawking effect due to Mathur [11], which is worth mentioning since it is quite often used in the literature. This analysis relies on a foliation of the black hole geometry into a set of "nice slices", as depicted on Fig. 3a. The slices extend both in the interior and the exterior of the black hole, and are chosen in such a way that they lie in a region with small curvature. The Hawking effect is then understood in the following pictorial way: notice that there is a stretching at the horizon between an initial slice and a later one, i.e. the late slice is longer than the initial one. As a consequence, a given wavemode with positive frequency is stretched when propagated from the initial slice to the late one as shown on Fig. 3b. On the late slice, it is thus no longer positive frequency. If with respect to the initial slice we are in the vacuum state, this will no longer be the case on the late slice: the quantum state corresponds on the late slice to a pair of entangled Hawking modes. As time passes, new Hawking pairs are created at the horizon. See [4, 11] for a more detailed discussion.



(a) Set of nice slices (b) Stretching of a wavemode

Figure 3: The nice slice picture : a wavemode is distorted when propagated from an initial spacelike slice to a later slice. This results in the creation of Hawking pairs (figures taken from [4]).

2.4 The transplanckian problem

We now briefly discuss how robust Hawking's result is. In the region around the horizon, the curvature of spacetime is low, so we expect quantum field theory in curved spacetime to be valid. However, there may be a problem related to the infinite redshift at the horizon. The outgoing modes are indeed infinitely blueshifted when propagated backwards and become transplanckian. It thus seems that the outgoing modes emerge from a "reservoir" of modes beyond the Planck length. It is clearly problematic that the properties of the Hawking radiation may depend on the existence of such transplanckian modes.

In his original derivation, Hawking propagated the outgoing modes backwards from future infinity back to past infinity through the collapsing matter. The problem with this reasoning is that some modes much below the Planck length at future infinity are blueshifted to unphysical transplanckian frequencies at past infinity. For this reason, in the derivation which I presented, we used the Cauchy slices Σ_i and Σ_f to avoid propagating the modes all the way back to infinity. Still, this derivation is leaving some room for doubt. The crucial assumption that we used to find the quantum state $|U\rangle$ was the local vacuum condition for short wavelength modes for the freely falling observer. This seems reasonable, since we can expect the modes with frequency much higher than $1/2m$ to be unexcited by the collapse process. However, if we want to evolve the quantum field starting from the initial state before to collapse to really determine whether those modes are excited, we cannot avoid involving transplanckian physics.

This question of whether the existence of the Hawking effect is dependent on short-distance behaviour has been investigated by many authors. In particular, analogue gravity settings which present a physical cutoff scale have provided some insight. The most famous of such models is the sonic black hole or "dumb hole" imagined by Unruh [12]. It consists of a fluid, whose velocity is larger in some region than the speed of sound, creating a sonic horizon. In this model, the intermolecular spacing provides a physical cutoff, and the dispersion relation of the waves is modified at short distances. Nevertheless, Unruh found using numerical simulations that Hawking radiation is present: there is no equivalent of a transplanckian reservoir, but the outgoing modes are instead produced from a "conversion" from the ingoing modes. Those models suggest that Hawking's result is quite robust to short-distance physics, nevertheless the question is not fully settled. See [13] for a review of the transplanckian question.

3 The information problem

The Hawking effect has the remarkable implication that black holes evaporate. Indeed the Hawking radiation carries energy to infinity, which by energy conservation implies that the black hole loses mass. In this section, I will discuss how the semiclassical approach leads us to expect that black holes completely evaporate and how this leads to information loss. I will then present the main alternatives about the fate of an evaporating black hole that are usually considered to possibly avoid this puzzling conclusion.

3.1 Black hole evaporation

In the framework of quantum field theory in curved spacetime, black hole evaporation appears as a consequence of the *backreaction* of the quantum field on the classical geometry. In our previous analyses, the spacetime metric has been kept fixed. In a semiclassical treatment in which the effects of "quantum fluctuations" are small and spacetime geometry can be treated classically, one can expect the quantum field to backreact via its energy-momentum tensor $T_{\mu\nu}$ on the classical metric. The most natural way is to describe backreaction by the "semiclassical Einstein equation"

$$G_{\mu\nu} = 8\pi\langle T_{\mu\nu}\rangle, \tag{72}$$

where $\langle T_{\mu\nu}\rangle$ is the expectation value of the energy-momentum tensor. Solving this means to find both a spacetime geometry $g_{\mu\nu}$ and a quantum state $|\Psi\rangle$ in this spacetime that satisfy (72). However, this approach suffers from serious difficulties. First, there are normal ordering ambiguities that make the definition of the energy-momentum tensor difficult in curved spacetime. Second, due to the presence of the energy-momentum tensor on the right hand side, the semiclassical Einstein equation has higher derivative terms than the usual Einstein equation, so it is likely to have solutions that blow up in

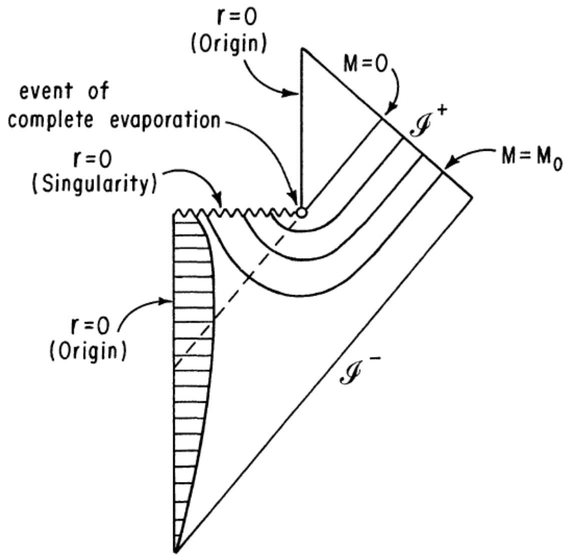


Figure 4: Penrose diagram of black hole evaporation. The contours drawn, starting from "$M = M_0$" to "$M = 0$", delimit the regions inside which the spacetime is approximately described by a Schwarzschild geometry with a mass M, which decreases with time from M_0 to 0 (figure taken from [9]).

a short time. Third, it is in practice extremely difficult to compute $\langle T_{\mu\nu} \rangle$ even for simple states. Therefore, solving self-consistently the backreaction is to difficult to be tractable even for simplified models, and has never been done.

Nevertheless, it is possible to estimate the lifetime of a black hole before its complete evaporation. For dimensional reasons the energy flux emitted by a black hole of mass M must be of the form

$$F = \frac{\alpha}{M^2}, \tag{73}$$

with α a dimensionless constant. As long as M is much larger that the Planck mass, we physically expect the black hole to be described by a locally Schwarzschild geometry with slowly decreasing mass M. The evolution of M will be given by

$$\frac{dM}{dt} = -F = -\frac{\alpha}{M^2}, \tag{74}$$

from which we get $M(t) = (M_0 - 3\alpha t)^{1/3}$. This implies that the black hole will evaporate or at least reach a Planckian size in a finite time, proportional to ${M_0}^3$.

This picture of black hole evaporation suggests the Penrose diagram depicted in Fig. 4. The way in which the mass of the black hole decreases is represented on the figure: the contours drawn separate the regions in which the geometry is close to the Schwarzschild geometry with a given mass M.

3.2 Information loss

The fact that black holes evaporate has a striking consequence: the semiclassical analysis implies that there is "information loss" after the black has evaporated completely. Let us explain what this precisely means.

Let us first consider a Cauchy slice Σ in the classical black hole spacetime where evaporation is not taken into account, as shown on Fig. 5. In our previous analysis, we showed that the restriction of the state of the quantum field on the part Σ_{out} of Σ which lies outside the black hole is a *mixed* state. This occurs because the modes outside the black hole are *entangled* with the modes inside the black hole.

Now, consider the black hole spacetime of Fig. 4, that we expect when evaporation is taken into account. The difference with the non-evaporating case is that after the black hole has evaporated the

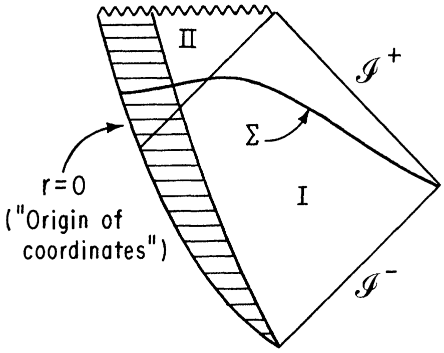


Figure 5: Spacetime diagram of a non-evaporating black hole, with a Cauchy slice Σ (figure taken from [9]).

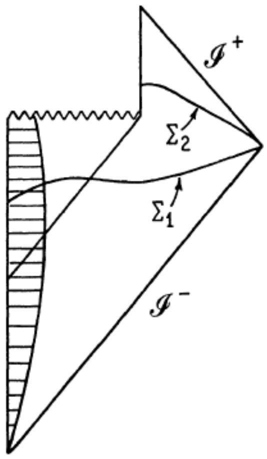


Figure 6: Spacetime diagram of an evaporating black hole, with two spacelike slices Σ_1 and Σ_2. The late slice Σ_2 is not a Cauchy slice, and the quantum state on Σ_2 is mixed (figure taken from [9]).

"entire state" of the field is now mixed. More precisely, let us consider two spacelike surfaces Σ_1 and Σ_2 as shown on Fig 6, where Σ_1 is located before the endpoint of the evaporation, and Σ_2 is located after the black hole has disappeared. The crucial point is that Σ_2 is *not* a Cauchy surface, indeed its past does not contain the black hole region. Therefore, the quantum state on Σ_2 is a *mixed*, and the evolution from Σ_1 to Σ_2 sends a pure initial state to a mixed state. This is often referred to as "information loss".

To understand well what is going on, it is important to ask ourselves what are the essential features of the semiclassical analysis that lead to this phenomenon of information loss. The first crucial point is that the interior and the exterior of the black hole are entangled. This entanglement is a robust conclusion, indeed it only relies on quantum field theory on the weakly curved spacetime region around the horizon. This point has been eloquently emphasized by Mathur in [4, 11], which I recommend reading. More generally, this is a feature of quantum field theory that different spacetime regions are entangled in a generic quantum state. The second important point is that the spacetime of Fig. 4 that we expect from the semiclassical reasoning is not globally hyperbolic. Intuitively, looking at the Penrose diagram, the reason is that because of the existence of the singularity, future infinity cannot extend "behind" the singularity, so there is an event horizon. This prevents the late slice Σ_2 after evaporation to be a Cauchy slice, so the Hawking radiation is in a mixed state after the black hole has evaporated completely.

3.3 The main alternatives

The semiclassical analysis thus strongly suggests that information is lost when a black hole evaporates. However, this conclusion is intriguing. Indeed in usual quantum mechanics, the evolution of a system is described by a unitary operator and sends pure states to pure states, so "information is not lost". The semiclassical picture of black hole evaporation is not expected to be the "ultimate" quantum gravity description with the "true" quantum gravity microstates, since spacetime is treated classically, so there may actually be corrections to that picture that could allow information not to be lost. If we trust the semiclassical picture in regions where the curvature is low, natural places to look for departures from the semiclassical theory are close to the singularity, or at the end stages of the evaporation, where the black hole reaches the Planck scale. For example the evaporation might stop at some time and we would be left with some kind of remnant. It has also been proposed that small corrections or even large departures from the semiclassical picture could allow information to get out of the black hole by some mechanism.

The main envisaged scenarios, as usually presented in the literature, are the following ones:

- Information is lost, as suggested by the semiclassical analysis.
- The black hole doesn't evaporate completely, there is a Planck-sized remnant which contains the "information" that was inside the black hole.
- The entanglement between modes inside and outside the black hole is transferred to the outside by some mechanism. The "information" is carried out by the Hawking radiation, there is no information loss.

Actually, there have been hundreds of papers published on the subject since Hawking's original publication forty years ago, and dozens of possibilities have been imagined for the fate of black holes and information. However, no consensus has been reached and the recent literature on the question is quite a big mess. The reason for this state of affairs is that, for the problem to be settled, we would need a quantum theory of gravity formulated precisely enough to describe the evaporation of a black hole. Such a theory would have both to tell what the fundamental quantum degrees of freedom and evolution laws are, and to make contact with the semiclassical picture which should be recovered in some limit. Unfortunately we are far from having such a theory at our disposal. As a result, the arguments that are put forward are often quite vague. From reading the recent literature on the subject, it clearly appears that the different experts have very different ideas about the resolution of the problem, and have a hard time convincing their colleagues of the appeal of their proposals.

For a non-expert trying to learn about the subject, my own experience shows that, when reading some paper or listening to some talk, it is difficult to grasp the global perspective of the author on the problem and how it articulates with other proposals. For this reason, I will try to review the main lines of thought about the information problem, emphasizing the inner logic and the essential assumptions of each proposal, and to analyze the consistence of the different pictures. I will begin by discussing the ideas related the the third possibility, i.e. that information is carried out by the Hawking radiation, which has become the opinion shared by most researchers. I will then review some other scenarios.

4 Insisting on keeping unitarity

It is now believed by most of the experts on the subject that the evaporation of a black hole is unitary and the information comes out in the Hawking radiation. Hawking himself has changed his mind and believes that information is not lost.

There are several motivations for believing in this picture, although it contradicts the one derived from the semiclassical analysis. One motivation is the idea that, in quantum gravity, black holes should be described as any quantum mechanical objects. In particular, there should be processes involving the creation and evaporation of virtual black holes. In the spirit of particle physics, one would like to be able to describe the formation and evaporation of a black hole in some S-matrix framework.

Furthermore, it has been argued in the influential paper [14] that information loss is incompatible with energy conservation. This point is still debated, but historically it lead many people to look for scenarios in which the evaporation is unitary. Another motivation is given by counting black hole microstates. The formula for the black hole entropy $S_{BH} = A/4$ suggests that in quantum gravity a black hole has a number of microstates given by $e^{S_{BH}}$. This number is decreasing when the black evaporates, so it seems that when the black hole becomes small, it cannot contain anymore all the information that fell into it and must start to release it. This point of view has been comforted by the microstate counting calculations in string theory which find the correct black hole entropy. A further motivation comes from AdS/CFT. The AdS/CFT conjecture indeed states that a theory of quantum gravity is the bulk is equal to a CFT without gravity on the boundary. The argument is then that in the boundary CFT the evolution is unitary, so in the bulk the evolution will be unitary as well and there is no information loss.

Motivated by those ideas, many proposals then *assume* that black hole evaporation is unitary and try to find some way for the information to come out. However, I am surprised at the way in which a lot of the proponents of unitary evaporation want this to be done. It would seem natural to me that if unitarity is to be "saved", it would be in the framework of a fully quantum theory of gravity, in which the gravitational degrees of freedom would be quantized, unlike in the semiclassical description. In such a theory, the spacetime structure may be different from the one of Fig. 4: maybe there will be no singularity, maybe there will be no event horizon, and everything will be unitary. To be satisfactory, the theory should give back the semiclassical description in some limit.

On the contrary, many people want to have a unitary evaporation while keeping the semiclassical geometry, and to have the information carried out by the Hawking radiation. This requires the entanglement of the outgoing modes with their ingoing partners to be "transferred" to the outgoing radiation by some mechanism. However, this is in sharp contradiction with the semiclassical description, in which the entanglement between the inside and the outside is robust to small corrections, in low curvature regions where we expect it to be valid. Having information conservation via the Hawking radiation thus implies large departures from the semiclassical picture and leads to new difficulties, as illustrated by the recent firewall controversy. In this section, I will discuss the main ideas that appear in this line of thought, including complementarity and firewalls. Recent review articles on the information problem written from this perspective are [5–7].

4.1 The Page curve

Let us thus for some time assume that information comes out in the Hawking radiation, follow the logic of this scenario, and describe the picture of black hole evaporation that it gives. On the spacetime diagram shown on Fig. 6, this means that we assume that the quantum state on Σ_2 is pure. To see how the information comes out, we can consider as argued by Page [15] the curve representing the von Neumann entropy of the Hawking radiation as a function of time (i.e. more precisely the von Neumann entropy of the restriction of full quantum state on the slices of constant Killing time outside the black hole), which is also the entanglement entropy S_{ent} between the inside and the outside. At the beginning of the evaporation, we expect the entropy of the radiation to increase because of the entanglement between the outgoing modes and their outgoing partners. At later times, we expect the entropy of the radiation to decrease and to reach zero at the end of the evaporation. The so-called *Page curve* will thus first rise, then go down, as illustrated on Fig. 7. The idea that the number of microstates inside the black hole is given by $e^{S_{BH}}$ gives another argument for this behaviour. Indeed, it implies that $S_{ent} \leq S_{BH}$, and S_{BH} decreases when the black hole evaporates. We thus expect that S_{ent} stops increasing at the *Page time*, when it becomes equal to S_{BH} and decreases from then on.

The Page curve is very different from the curve that is predicted by the semiclassical analysis, as emphasized by Mathur [4]. Indeed, in the "nice slice" picture of [4], new entangled Hawking modes are created each time we go from one slice to the next slice. As a consequence, the entropy of the Hawking radiation increases with time at a steady rate until the end of the evaporation, as shown on Fig. 7 by the dotted line. The increase of the entanglement entropy is insensitive to small corrections to the quantum state of the field. The proposal that the information comes out with the Hawking radiation is therefore a strong violation of the semiclassical picture.

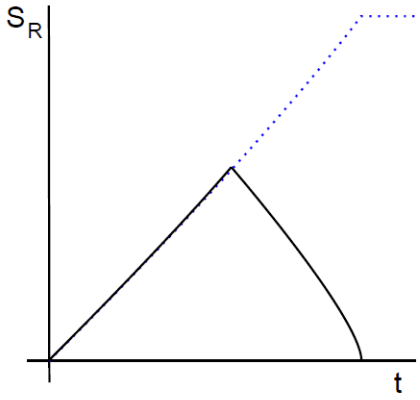


Figure 7: The Page curve represents the von Neumann entropy of the radiation S_R as a function of time: the entropy first increases, then goes down to zero. The dotted line is the "Hawking curve" predicted by the semiclassical analysis: the entropy of the radiation increases until the end of the evaporation (figure taken from [7]).

4.2 Complementarity and firewalls

In this line of thought in which information is carried out by the Hawking radiation, it is often imagined that there is some kind of effective membrane at a Planck distance from the horizon, sometimes referred to as a *streched horizon*, which retains the information for some time and then releases it. This is for example the case in 't Hooft's "brick wall model" [16]. But this picture is not very realistic since it doesn't involve the black hole interior and doesn't tell what happens for an observer falling into the black hole.

A proposal named *black hole complementarity*, in analogy with complementarity in quantum mechanics, answers the problem by saying that the information is somehow *both* outside and inside the black hole [17]. This would be problematic if a single observer could see both "copies" of the information, which would constitute quantum cloning forbidden by the no-cloning theorem [18]. The idea of complementarity is that this would never happen: an observer could never see the two copies of the information, so there would be no problem.

Notice that this scenario is clearly incompatible with the semiclassical picture in which there is no possibility for the quantum state to be "duplicated" from one Cauchy surface to a later Cauchy surface. This means that if one believes in complementarity, then the general framework of the analysis has to be drastically changed. For example, speaking of the quantum state on a spatial slice that is not entirely contained in the past of a single observer should have no meaning in the new framework. Another idea, called strong complementarity, is that there is no global Hilbert space, but instead each observer has its own Hilbert space, with matching conditions when they overlap. The proponents of black hole complementarity formulated postulates that summarize the picture expected from this scenario:

1. **Unitarity.** The process of formation and evaporation of a black hole is unitary as seen from a distant observer, in particluar there exists a S-matrix which describes the evolution from the infalling matter to the outgoing Hawking radiation.

2. **Effective field theory holds.** Outside the stretched horizon of a black hole, physics can be well described by the semiclassical theory, i.e. by quantum fields on a classical background.

3. **Microscopic interpretation of the Bekenstein-Hawking entropy.** To a distant observer a black hole appears as an object with discrete energy level. The subspace of states describing a black hole of mass M is given by the exponential of the corresponding Bekenstein-Hawking entropy.

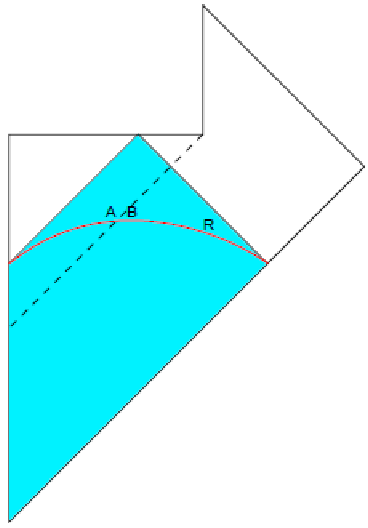


Figure 8: A Cauchy slice contained in the past of a single observer, split into an outgoing mode B, its ingoing partner A and the rest of the radiation R. If we assume both that the information comes out and that the Hawking partners are entangled as predicted by the semiclassical analysis, then B is entangled both with A and R, which is forbidden by the strong subbaditivity of the von Neumann entropy (figure taken from [7]).

4. **No drama.** An infalling observer experiences nothing particular when he crosses the horizon. In particular he sees no high-energy excitations.

For some time, black hole complementarity has been accepted by many researchers. However, this picture rises new difficulties. Putting together some previous arguments, the authors of [19], usually referred to as AMPS, argued that the postulates of black hole complementarity are inconsistent (actually only postulates 1, 2 and 4, postulate 3 does not enter the argument). These authors suggest that it is the postulate 4 that should be abandoned: an infalling observer should meet high-energy particles when crossing the horizon, i.e. there would be a "firewall" at the horizon. The firewall argument created a lot of debate in community. It was followed by dozens of papers presenting arguments for or against the existence of firewalls, and generated many new ideas.

Before going into some more detail, let us give a sketch of the AMPS argument. On the one hand, if an infalling observer sees no incoming excitations, this implies that an outgoing mode B is entangled with its ingoing partner A, as we have seen in our analysis of the Hawking effect. On the other hand, if the information comes out in the Hawking radiation, this means that the mode B must be entangled with the outgoing modes emitted earlier. But to have B entangled both with A and the "early radiation" is forbidden by a general property of entanglement in quantum mechanics called *monogamy of entanglement*, which roughly tells that a system cannot be both entangled with two other systems. We thus have a contradiction. If we abandon the entanglement between B and A, then the quantum state cannot be smooth at the horizon and we have a firewall.

Let us see more precisely how this is formulated. The setting is shown on Fig. 8. The Cauchy slice on the figure is split into three segments corresponding to the mode B, its partner A and the radiation emitted earlier R. The key point is that the slice is entirely contained in the past of a single observer, so we cannot invoke complementarity to escape the problem. The argument then invloves inequalities between different von Neumann entropies. From a general property of the von Neumann entropy called *strong subbaditivity* (see e.g. [20]), we have the inequality

$$S_{RB} + S_{BA} \geq S_B + S_{ABR}, \tag{75}$$

where S_{RB} denotes the von Neumann entropy of the restricted density matrix on BR, and so on. Now, if we are at late times and assume that the entropy of the radiation is decreasing with time according to the Page curve, we have $S_{BR} < S_R$. Furthermore if we assume there is entanglement across the horizon as predicted by the semiclassical calculation, the system AB is in the pure Unruh vacuum, so $S_{AB} = 0$ and $S_{ABR} = S_R$. Combining those three relations, we get $S_A \geq S_B + S_A$, hence $S_B = 0$. Since the restricted quantum state on B is mixed (we have seen it is given by a thermal density matrix), we have a contradiction.

The firewall controversy has triggered a revival of the black hole information paradox, and has given rise to many arguments for or against firewalls. Among the current lines of discussion, let us mention the question of whether AdS/CFT tells us about the black hole interior, the idea proposed in [21] that it would take to much time to do the quantum computation necessary to measure the entanglement, possible stronger versions of complementarity, or the so-called $ER = EPR$ conjecture which relates quantum entanglement and transversable wormholes. See [5–7] for further discussion and more comprehensive references on the post-firewall literature.

Until now in this section, I have tried to describe the unitary evaporation paradigm following its own logic. I will now discuss the validity of the assumptions made in this scenario. Actually, I am not much convinced by the picture that unitary evaporation proposes, mainly on logical grounds. Indeed, it seems to me that the logic of this picture is somehow hybrid: at the same time, some ingredients like the Page curve drastically differ from the semiclassical picture, but some of the semiclassical ingredients are conserved, for example in the AMPS argument, the quantum degrees of freedom are the Hawking modes and the Hilbert space factorizes in the same way as in the semiclassical description. More generally the arguments in the literature are often quite vague. Indeed, unlike the semiclassical description, the unitary evaporation picture relies on no well-defined and precise framework. For example it is worth noticing how the black hole complementarity idea is formulated: the features of the evaporation that one might expect are expressed under the form of postulates, but there is no detailed enough proposal for how they can be implemented. Similarly, it is often claimed that AdS/CFT implies that evaporation is unitary. But, as emphasized e.g. in [4], this conclusion is only drawn on general grounds and doesn't tell where the semiclassical picture fails. The proponents of unitary evaporation of course acknowledge this kind and criticism and agree that their scenarios have to be made more precise, but they still believe they will be able to come up with a consistent picture. I am more worried than they are about the lack of a well-defined framework: it seems to me that this strongly threatens the very relevance of the arguments. In this spirit, the firewall paradox appears as a consequence of the inconsistency of the premises of the discussion. The first postulate which tells that the information goes out in the Hawking radiation is inconsistent with the semiclassical picture, and this is the one that should be abandoned.

5 Other scenarios

I will now discuss other approaches to the black hole information problem that are logically distinct from what I have called the unitary evaporation paradigm and are less radical in the sense that they take the semiclassical picture more seriously.

5.1 The fuzzball proposal

One approach that has been investigated a lot is the fuzzball proposal [22] initiated by Mathur. Its proponents claim that they have solved the information problem. The fuzzball scenario is formulated within string theory and proposes that there are actually no black holes. Instead, a cloud of collapsing matter tunnels to a configuration of "fuzzy" stringy microstates involving D-branes, which extend in the whole interior of the would-be black hole and present neither a horizon nor a singularity.

The logic which motivates this picture is clearly formulated by Mathur in [4,11]. He emphasizes the importance of the semiclassical picture, which he rephrases as the "Hawking theorem": if we assume that there are no quantum gravity effects in low-curvature regions i.e. if (i) in such regions the physics is well described by quantum field theory in weakly curved spacetime and (ii) traditional black holes with a smooth horizon form in the theory, then there will be information loss or remnants. According to Mathur, many proposals about black hole information, including the claim that AdS/CFT implies unitarity, are just irrelevant since they don't tell where the hypotheses of the Hawking theorem fail.

The fuzzball proposal says that the hypotheses of the Hawking theorem are actually not satisfied, i.e. that quantum gravity effects manifest themselves at a distance of the order of the black hole size. Let us briefly describe the main features of the fuzzball picture. The fact that stringy degrees of freedom can extend to such big distances crucially relies on the existence of compactified extra dimensions of spacetime on which they can propagate. The geometry of these microstates presents no singularity and no event horizon. The collapsing matter has a small probability of order e^{-S} to tunnel to one given microstate, but the number of microstates is of order e^S, so the total tunnelling probability can be of order $O(1)$. Furthermore, an estimate shows that the tunnelling time is much shorter than the evaporation time, so the infalling matter will indeed have time to tunnel into a fuzzball. The crucial point in the fuzzball scenario, which allows quantum gravity effects to appear although the curvature is small, is the huge number e^S of microstates, and much of the activity in this program is to find enough different microstates to account for the black hole entropy.

However, the fuzzball proposal has limits. It assumes that string theory is the correct theory of quantum gravity, which may not be the case, it works with extra dimensions and involves supersymmetric near-extremal black holes which are far from the physical ones. The question of what happens to an infalling observer is also unclear. As far as the logic is concerned, I find the fuzzball proposal more satisfactory than the unitary evaporation paradigm, since it is formulated in a quite well-defined framework and doesn't break the logic of the semiclassical picture. Nevertheless it is a radical proposal in the sense that it affirms that quantum gravity effects manifest themselves at places we would otherwise expect the semiclassical analysis to be valid.

5.2 Conservative resolutions

I will now discuss other scenarios which can be described as conservative since they respect the semiclassical picture. The most natural way for unitarity to be possibly restored is in a full quantum theory of gravity in which the gravitational degrees of freedom are quantized and the semiclassical description is obtained in some limit. In such a theory there may be no singularity and/or no horizon in the full quantum description. The authors of [23] classified conservative resolutions according to whether they present an event horizon or a singularity. Hawking's original proposal that information is lost and the remnant or baby universe scenarios fit in that scheme: the first presents both a singularity and an event horizon and the second an event horizon but no singularity.

The remnant scenario is not a radical proposal, since it only requires departures from the semiclassical analysis at the end of the evaporation. Nevertheless there are several difficulties with this scenario. To be entangled with the Hawking radiation in such a way that the total state be pure, a Planck-sized remnant should have an extremly high or even infinite density of states. But then the production of remnants would entropically infinitely favoured over any other thermodynamical process. Nevertheless, it has been pointed out [23] that this standard criticism of remnants is not logically decisive. For example a remnant might have a "bag of gold" geometry with a Planckian area but a huge internal volume. Even if the case for or against remnants is not fully settled, I am aware of no current advocates of remnants. See [24] for a recent review of the different remnant scenarios.

Another possibility is that the full quantum gravity solution contains no event horizon and no singularity. This is actually the picture suggested by loop quantum gravity, according to which the singularity is resolved in some simple models. Interestingly, a two-dimensional model of quantum gravity, in which explicit calculations of the black hole evaporation can be done and which seems to present those features, has been studied in [25]. In this model, the Penrose diagram is not the one of Fig 4 predicted by the semiclassical analysis: the would-be singularity is replaced by a region with high quantum fluctuations of the geometry, future infinity is "longer" and extends behind the would-be singularity. The evolution between past infinity and the longer future infinity is unitary. It seems to me that such a kind of resolution is what makes more sense, since it locally agrees with the semiclassical picture in low curvature regions where we expect it to be valid, and the "quantum gravity effects" only appear close to the would-be singularity, as shown on Fig. 9. However, the model of [25] does not describe a physical black hole, and no detailed enough quantum gravity model of a four-dimensional black hole along the same lines is known. I think it would be very interesting to investigate further such models.

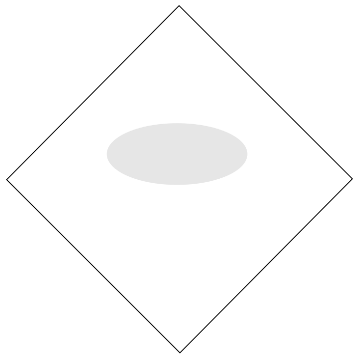


Figure 9: Penrose diagram suggested by [25]: the would-be singularity is replaced by a region with high quantum fluctuations of the geometry, and future infinity extends behind the would-be singularity.

In the absence of a satisfactory fully quantum gravitational description, the most reasonable attitude is perhaps to stick to information loss. This is the opinion of many classical relativists who unlike Hawking didn't convert to unitary evaporation [26].

Conclusion

The phenomenon of black hole evaporation presents a tension between general relativity and quantum mechanics. This arises because quantum field theory in curved spacetime can only be formulated in a globally hyperbolic region. In the semiclassical picture the expected spacetime geometry is not globally hyperbolic, and this implies that we have an evolution from pure to mixed states, unlike in ordinary quantum mechanics. Despite the hundreds of different ideas that have been proposed to solve the information problem since it was formulated 40 years ago and the thousands of papers which were written on the subject, it is fair to say that the opinion of the community has not thermalized to a consensus.

Very different perspectives

Now that we have discussed the main lines of thought, it is possible to appreciate how big the differences are between the opinion of the experts. It seems to me that these sharp divergences depend on which fundamental principles one wants to hold on most. Does one above all want to have everything described in a unitary quantum mechanical framework ? Or does one want the semiclassical picture to be valid close to the horizon ?

As an example of incompatible perspectives, the Bekenstein-Hawking entropy can be interpreted in very different ways, and the interpretation one chooses has a strong influence on the way one thinks about the information paradox [27]. The proponents of the unitary evaporation paradigm believe that the Bekenstein-Hawking entropy counts the number of microstates inside the volume of the black hole. This suggests that as evaporation progresses the black hole has not enough states to contain the information, so the information must leak out in the Hawking radiation. However, other researchers, among them those working in loop quantum gravity, have the exactly opposite perspective. In this picture, the information cannot go out since the outside is causally separated from the inside of the black hole, and the fact that a black hole evaporates *shows* that the strong interpretation of black hole entropy cannot hold. Instead, the weak interpretation of black hole entropy states that the entropy is a property of the horizon: it counts the number of freedom that can interact at the horizon. In the weak interpretation, there is no problem of principle for a black hole to have an arbitrarily large number of states. Another example is the argument using strong subadditivity. Mathur formulated this argument in [11] to show that the entropy of the Hawking radiation always increases during evaporation, so that the black hole does *not* follow the Page curve. A couple of years later, AMPS *assumed* that a black hole follows the Page curve and used the subadditivity argument to conclude that we must have a firewall, as we saw in section 4.

To realise how divergent the opinion of the experts are, it is amusing to do a short literary analysis of a few recent review articles [5, 6, 11, 26] written from very different perspectives.

Joseph Polchinski The main thread of Polchinski's recent review [6] on the information paradox is AdS/CFT. "One could go around in circles" between the three classical alternatives, he ironically says, but fortunately AdS/CFT comes to save us and to tell us that information is not lost. Then, asking where Hawking's calculation fails leads us to the possibility of firewalls and raises new open questions. Polchinski says he is "agnostic" between the different possibilities.

Don Marolf His recent short review [5] is built on the strong interpretation of the Bekenstein-Hawking entropy. Marolf calls the information problem which arises at the end of the evaporation a "straw man information problem", since it can easily be evaded by quantum gravity corrections at the Planck scale. In his opinion, the serious information problem occurs long before the end of evaporation, at the Page time. This is an argument we have already discussed: the strong interpretation of the Bekenstein-Hawking entropy implies that an evaporating black hole follows the Page curve, which is in conflict with the semiclassical picture.

Samir Mathur His opinions about the information problem are very clearly expressed in several papers [4, 11, 28]. The punchline of those articles is that any proposal that claims to solve the information paradox should address the semiclassical logic that he reformulates as the "Hawking theorem", and precisely tell where it fails. He complains that "the black hole information paradox is a very poorly understood problem", and that many arguments are "either incorrect or irrelevant". In particular, the statement that AdS/CFT implies that information is not list is "completely circular". Mathur claims that the fuzzball proposal solves the information paradox. It addresses the Hawking theorem by showing that its hypotheses are not satisfied: quantum gravity effects actually manifest themselves on large distances.

Robert Wald He expresses his views on the information problem in a recent article [26] written together with Unruh. It is interesting to notice that, despite the hundreds of papers that have been written in the meantime, his opinion is exactly the same that the one he expressed more than 20 years ago in his book [9]. Wald and Unruh trust the semiclassical description: "We remain firm in our belief in the validity of quantum theory in regimes away from the Planck scale", and think that information is lost as predicted by the semiclassical analysis, the black hole acting as a sink for the ingoing radiation. They insist that the semiclassical picture respects the *local* evolution laws and are critical against the proposals that violate this picture to ensure *global* unitary evolution: "We find it ironic that some researchers who may have been seeking to 'save quantum mechanics' by trying to evade the semiclassical arguments that a pure state evolves to a mixed state are, in fact, effectively attempting to 'destroy quantum mechanics' by seeking to modify quantum mechanics".

What do I think ?

It is quite perplexing to see how incompatible the views of such undoubtedly clever and serious people are, and doesn't help to know what to think about all that. I think that the honest answer is that nobody knows what ultimately happens when a black hole evaporates. It would seem presumptuous to formulate a personal opinion, given that I am not an expert and researchers who have worked on the subject for decades fail to agree. However, looking at things differently, this very lack of consensus can encourage me to give an opinion. After all, whatever I may say, there will be somebody who strongly disagrees, so let's not be shy!

Which opinion did I get from this study of the information paradox? Since we have no satisfactory enough theory of quantum gravity and probably won't for some time, it seems important to me to think on logical grounds and to draw a picture of black hole evaporation that is self-consistent. In particular, it doesn't make much sense to formulate arguments outside a well-defined framework. The semiclassical framework inside which the Hawking effect is derived is well-defined, and we can safely expect it to be valid to a good approximation in regions where the curvature is small. The semiclassical picture suggests that the spacetime geometry is the one of Fig. 4, which implies that information is lost after the black hole has evaporated. One shouldn't worry not to have a unitary evolution in this picture, since we don't expect the semiclassical framework to be the "true" theory of quantum gravity, but rather to be obtained as a limit of a more fundamental fully quantum theory of gravity which could be unitary, with different fundamental degrees of freedom. There is absolutely no problem in having both a unitary evolution in the fundamental "microscopic" theory and information loss in the semiclassical "macroscopic" limit. Indeed, in such a picture, the "information" about the infalling matter would be transferred to the quantum gravitational degrees of freedom. From the coarse-grained semiclassical perspective, those degrees of freedom are "hidden" and information is effectively lost. I am actually surprised that the dominant paradigm, according to which the information comes out in the Hawking radiation, is somehow logically "hybrid" since it keeps the degrees of freedom of the semiclassical picture (i.e. the Hawking modes) while drastically violating quantum field theory in curved spacetime at places it should be valid. It makes more sense to think that the information is stored in correlations with new hidden gravitational degrees of freedom rather than in "subtle correlations" in the Hawking radiation. This picture is actually close to the one motivated by loop quantum gravity [29].

Such a picture is actually very similar to what happens in the statistical mechanics of an ideal gas. At the microscopic level, an ideal gas is described by the positions and velocities of all the molecules. The evolution of the microscopic degrees of freedom is totally deterministic and reversible, and "information" is never lost. However, in practice we never have access to the microscopic degrees of freedom, and we have to introduce a coarse-grained description. For the perfect gas, we can use a mesoscopic description in which the variables that describe the system are the distributions of particles in the phase space of positions and velocities. From the dynamics of the microscopic variables, one can derive the dynamics of these new variables: it is given by the Boltzmann equation. Now something remarkable happens: the mesoscopic dynamics is irreversible, as expressed by Boltzmann's H-theorem, i.e. information is lost! What happens is that the initial information has leaked to the inaccessible microscopic degrees of freedom, as explained in an illuminating way in [30]. At a further coarse-grained level, the gas is described by thermodynamical variables, and the macroscopic evolution is irreversible as expressed by the growth of the thermodynamical entropy. I think that it makes sense to expect a similar picture for black hole mechanics. There is just a little problem: we don't know the microscopic theory.

The big question remains to be able to formulate a consistent and well-defined theory of quantum gravity with the right semiclassical limit, and it seems to me that the question of black hole evaporation does not significantly augment the challenge. Steps have been done in that direction, but we are currently far from having a precise enough framework to address black hole evaporation. Until we have such a nice theory, if it ever exists, and with the absence of any guidance from experiment, my feeling is that most arguments for or against information loss cannot but be quite vague and often irrelevant. Anyway, it seems that the characteristic time for the evolution of the subject is several decades, so it is likely that we have to wait for quite some time before we know the end of the story.

References

[1] S. W. Hawking. Particle Creation by Black Holes. *Commun. Math. Phys.*, 43:199–220, 1975.

[2] J. M. Bardeen, B. Carter, and S. W. Hawking. The four laws of black hole mechanics. *Comm. Math. Phys.*, 31(2):161–170, 1973.

[3] S. W. Hawking. Breakdown of Predictability in Gravitational Collapse. *Phys. Rev.*, D14:2460–2473, 1976.

[4] Samir D. Mathur. What Exactly is the Information Paradox? *Lect. Notes Phys.*, 769:3–48, 2009.

[5] Donald Marolf. The Black Hole information problem: past, present, and future. 2017.

[6] Joseph Polchinski. The Black Hole Information Problem. In *Proceedings, Theoretical Advanced Study Institute in Elementary Particle Physics: New Frontiers in Fields and Strings (TASI 2015): Boulder, CO, USA, June 1-26, 2015*, pages 353–397, 2017.

[7] Daniel Harlow. Jerusalem Lectures on Black Holes and Quantum Information. *Rev. Mod. Phys.*, 88:15002, 2016.

[8] Ted Jacobson. Introduction to quantum fields in curved space-time and the Hawking effect. In *Lectures on quantum gravity. Proceedings, School of Quantum Gravity, Valdivia, Chile, January 4-14, 2002*, pages 39–89, 2003.

[9] Robert M. Wald. *Quantum Field Theory in Curved Space-Time and Black Hole Thermodynamics*. Chicago Lectures in Physics. University of Chicago Press, Chicago, IL, 1995.

[10] Viatcheslav F Mukhanov and Sergei Winitzki. *Introduction to quantum effects in gravity*. Cambridge Univ. Press, Cambridge, 2007.

[11] Samir D. Mathur. The Information paradox: A Pedagogical introduction. *Class. Quant. Grav.*, 26:224001, 2009.

[12] W. G. Unruh. Sonic analogue of black holes and the effects of high frequencies on black hole evaporation. *Phys. Rev. D*, 51:2827–2838, Mar 1995.

[13] Ted Jacobson. Trans Planckian redshifts and the substance of the space-time river. *Prog. Theor. Phys. Suppl.*, 136:1–17, 1999.

[14] Tom Banks, Leonard Susskind, and Michael E. Peskin. Difficulties for the Evolution of Pure States Into Mixed States. *Nucl. Phys.*, B244:125–134, 1984.

[15] Don N. Page. Average entropy of a subsystem. *Phys. Rev. Lett.*, 71:1291–1294, 1993.

[16] Gerard 't Hooft. On the Quantum Structure of a Black Hole. *Nucl. Phys.*, B256:727–745, 1985.

[17] Leonard Susskind, Larus Thorlacius, and John Uglum. The Stretched horizon and black hole complementarity. *Phys. Rev.*, D48:3743–3761, 1993.

[18] W. K. Wootters and W. H. Zurek. A single quantum cannot be cloned. *Nature*, 299:802–803, 1982.

[19] Ahmed Almheiri, Donald Marolf, Joseph Polchinski, and James Sully. Black Holes: Complementarity or Firewalls? *JHEP*, 02:062, 2013.

[20] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, New York, NY, USA, 10th edition, 2011.

[21] Daniel Harlow and Patrick Hayden. Quantum Computation vs. Firewalls. *JHEP*, 06:085, 2013.

[22] Samir D. Mathur. The Fuzzball proposal for black holes: An Elementary review. *Fortsch. Phys.*, 53:793–827, 2005.

[23] Sabine Hossenfelder and Lee Smolin. Conservative solutions to the black hole information problem. *Phys. Rev.*, D81:064009, 2010.

[24] Pisin Chen, Yen Chin Ong, and Dong-han Yeom. Black Hole Remnants and the Information Loss Paradox. *Phys. Rept.*, 603:1–45, 2015.

[25] Abhay Ashtekar, Victor Taveras, and Madhavan Varadarajan. Information is Not Lost in the Evaporation of 2-dimensional Black Holes. *Phys. Rev. Lett.*, 100:211302, 2008.

[26] William G. Unruh and Robert M. Wald. Information Loss. 2017.

[27] Ted Jacobson, Donald Marolf, and Carlo Rovelli. Black hole entropy: Inside or out? *Int. J. Theor. Phys.*, 44:1807–1837, 2005.

[28] Samir D. Mathur. What the information paradox is not. 2011.

[29] Alejandro Perez. Black Holes in Loop Quantum Gravity. 2017.

[30] R. Balian. Information in statistical physics. *Studies in the History and Philosophy of Modern Physics*, 36:323–353, 2005.