1 Introduction

In this review we discuss some recent progress on aspects of the black hole information paradox.

Before delving into it let us discuss a big picture motivation. One of the main motivations to study quantum gravity is to understand the earliest moments of the universe, where we expect that quantum effects are dominant. In the search for this theory, it is better to consider simpler problems. A simpler problem involves black holes. They also contain a singularity in their interior. It is an anisotropic big crunch singularity, but it is also a situation where quantum gravity is necessary, making it difficult to analyze. Black holes, however, afford us the opportunity to study them as seen from the outside. This is simpler because far from the black hole we can neglect the effects of gravity and we can imagine asking sharp questions probing the black hole from far away. One of these questions will be the subject of this review. We hope that, by studying these questions, we will eventually understand the black hole singularity and learn some lessons for the big bang, but we will not do that here.

Studies of black holes in the '70s showed that black holes behave as thermal objects. They have a temperature that leads to Hawking radiation. They also have an entropy given by the area of the horizon. This suggested that, from the point of view of the outside, they could be viewed as an ordinary quantum system. Hawking objected to this idea through what we now know as the "Hawking information paradox." He argued that a black hole would destroy quantum information, and that the von Neumann entropy of the universe would increase by the process of black hole formation and evaporation. Results from the '90s using string theory, a theory of quantum gravity, provided some precise ways to study this problem for very specific gravity theories. These results strongly suggest that information does indeed come out. However, the current understanding requires certain dualities to quantum systems where the geometry of spacetime is not manifest.

During the past 15 years, a better understanding of the von Neumann entropy for gravitational systems was developed. The computation of the entropy involves also an area of a surface, but the surface is not the horizon. It is a surface that minimizes the generalized entropy. This formula is almost as simple as the Bekenstein formula for black hole entropy [1,2]. More recently, this formula was applied to the black hole information problem, giving a new way to compute the entropy of Hawking radiation [3,4]. The final result differs from Hawking's result and is consistent with unitary evolution.

The first version of the fine-grained entropy formula was discovered by Ryu and Takayanagi [5]. It was subsequently refined and generalized by a number of authors [3,4,6–11]. Originally, the Ryu-Takayanagi formula was proposed to calculate holographic entanglement entropy in anti-de Sitter spacetime, but the present understanding of the formula is much more general. It requires neither holography, nor entanglement, nor anti-de Sitter spacetime. Rather it is a general formula for the fine-grained entropy of quantum systems coupled to gravity.

Our objective is to review these results for people with minimal background in this problem. We will not follow a historical route but rather try to go directly to the final formulas and explain how to use them. For that reason, we will not discuss many related ideas that served as motivation, or that are also very useful for the general study of quantum aspects of black holes. A sampling of related work includes [12–35].

Many details and caveats will necessarily be swept under the rug, although we will discuss some potential technical issues in the discussion section. We believe the caveats are only technical and unlikely to change the basic picture.

2 Preliminaries

2.1 Black hole thermodynamics

When an object is dropped into a black hole, the black hole responds dynamically. The event horizon ripples briefly, and then quickly settles down to a new equilibrium at a larger radius. It was noticed in the 1970s that the resulting small changes in the black hole geometry are constrained by equations closely parallel to the laws of thermodynamics [1, 2, 36–42]. The equation governing the response of a rotating black hole is [40]

$$\frac{\kappa}{8\pi G_N} d\,(\text{Area}) = dM - \Omega dJ \ , \tag{2.1}$$

where κ is its surface gravity[1], M is its mass, J is its angular momentum, and Ω is the rotational velocity of the horizon. The area refers to the area of the event horizon, and G_N is Newton's constant. If we postulate that the black hole has temperature $T \propto \kappa$, and entropy $S_{\text{BH}} \propto$ Area, then this looks identical to the first law of thermodynamics in the form

$$T dS_{\text{BH}} = dM - \Omega dJ \ . \tag{2.2}$$

In addition, the area of the horizon always increases in the classical theory [38], suggesting a connection to the second law of thermodynamics. This is just a rewriting of the Einstein equations in suggestive notation, and initially, there was little reason to believe that it had anything to do with 'real' thermodynamics. In classical general relativity, black holes have neither a temperature nor any significant entropy. This changed with Hawking's discovery that, when general relativity is coupled to quantum field theory, black holes have a temperature [42]

$$T = \frac{\hbar\kappa}{2\pi} \ . \tag{2.3}$$

[1]Unfortunately, the name "surface gravity" is a bit misleading since the proper acceleration of an observer hovering at the horizon is infinite. κ is related to the force on a massless (unphysical) string at infinity, see e.g. [43].

(We set $c = k_B = 1$.) This formula for the temperature fixes the proportionality constant in $S_{\rm BH} \propto$ Area. The total entropy of a black hole and its environment also has a contribution from the quantum fields outside the horizon. This suggests that the total or 'generalized' entropy of a black hole is [2]

$$S_{\rm gen} = \frac{\text{Area of horizon}}{4\hbar G_N} + S_{\rm outside} \ , \tag{2.4}$$

where $S_{\rm outside}$ denotes the entropy of matter as well as gravitons outside the black hole, as it appears in the semiclassical description. It also includes a vacuum contribution from the quantum fields [44].[2] The generalized entropy, including this quantum term, is also found to obey the second law of thermodynamics [45],

$$\Delta S_{\rm gen} \geq 0 \ , \tag{2.5}$$

giving further evidence that it is really an entropy. This result is stronger than the classical area theorem because it also covers phenomena like Hawking radiation, when the area decreases but the generalized entropy increases due to the entropy of Hawking radiation.

The area is measured in Planck units, $l_p^2 = \hbar G_N$, so if this entropy has an origin in statistical mechanics then a black hole must have an enormous number of degrees of freedom. For example, the black hole at the center of the Milky Way, Sagittarius A*, has

$$S \approx 10^{85} \ . \tag{2.6}$$

Even for a black hole the size of a proton, $S \approx 10^{40}$. In classical general relativity, according to the no-hair theorem, there is just one black hole with mass M and angular momentum J, so the statistical entropy of a black hole is naively zero. Including quantum fields helps, but has not led to a successful accounting of the entropy. Finding explicitly the states giving rise to the entropy is an interesting problem, which we will not discuss in this review.

2.2 Hawking radiation

The metric of a Schwarzschild black hole is

$$ds^2 = -\left(1 - \frac{r_s}{r}\right) dt^2 + \frac{dr^2}{1 - \frac{r_s}{r}} + r^2 d\Omega_2^2 \ . \tag{2.7}$$

[2] The quantum contribution by itself has an ultraviolet divergence from the short distance entanglement of quantum fields across the horizon. This piece is proportional to the area, A/ϵ_{uv}^2. However, matter loops also lead to an infinite renormalization of Newton's constant, $1/(4G_N) \to \frac{1}{4G_N} - \frac{1}{\epsilon_{uv}^2}$. Then these two effects cancel each other so that $S_{\rm gen}$ is finite. As usual in effective theories, these formally "infinite" quantities are actually subleading when we remember that we should take a small cutoff but not too small, $l_p \ll \epsilon_{uv} \ll r_s$.

The Schwarzschild radius $r_s = 2G_N M$ sets the size of the black hole. We will ignore the angular directions $d\Omega_2^2$ which do not play much of a role. To zoom in on the event horizon, we change coordinates, $r \to r_s(1 + \frac{\rho^2}{4r_s^2})$, $t \to 2r_s\tau$, and expand for $\rho \ll r_s$. This gives the near-horizon metric

$$ds^2 \approx -\rho^2 d\tau^2 + d\rho^2 \ . \tag{2.8}$$

To this approximation, this is just flat Minkowski spacetime. To see this, define the new coordinates

$$x^0 = \rho \sinh \tau, \qquad x^1 = \rho \cosh \tau \tag{2.9}$$

in which

$$ds^2 \approx -\rho^2 d\tau^2 + d\rho^2 = -(dx^0)^2 + (dx^1)^2 \ . \tag{2.10}$$

Therefore according to a free-falling observer, the event horizon $r = r_s$ is not special. It is just like any other point in a smooth spacetime, and in particular, the geometry extends smoothly past the horizon into the black hole. This is a manifestation of the equivalence principle: free-falling observers do not feel the effect of gravity. Of course, an observer that crosses the horizon will not be able to send signals to the outside[3].

The spacetime geometry of a Schwarzschild black hole that forms by gravitational collapse is illustrated in fig. 1. An observer hovering near the event horizon at fixed r is accelerating — a rocket is required to avoid falling in. In the near-horizon coordinates (2.10), an observer at fixed ρ is following the trajectory of a uniformly accelerated observer in Minkowski spacetime.

A surprising fact is that a uniformly accelerating observer in flat space detects thermal radiation. This is known as the Unruh effect [46]. There is a simple trick to obtain the temperature [47]. The coordinate change (2.9) is very similar to the usual coordinate change from Cartesian coordinates to polar coordinates. It becomes identical if we perform the Wick rotation $\tau = i\theta$, $x^0 = ix_E^0$; then

$$x_E^0 = \rho \sin \theta, \quad x^1 = \rho \cos \theta \ . \tag{2.11}$$

The new coordinates (x_E^0, x^1) or (ρ, θ) are simply Cartesian or polar coordinates on the Euclidean plane $\mathbb{R}^2$. In Euclidean space, an observer at constant ρ moves in a circle of length $2\pi\rho$. Euclidean time evolution on a circle is related to the computation of thermodynamic quantities for the original physical system (we will return to this in section 2.3). Namely, $\mathrm{Tr}\,[e^{-\beta H}]$ is the partition function at temperature $T = 1/\beta$. β is the length of the Euclidean time evolution and the trace is related to the fact that we are on a circle. This suggests that

[3]We can say that the interior lies behind a Black Shield (or Schwarz Schild in German).

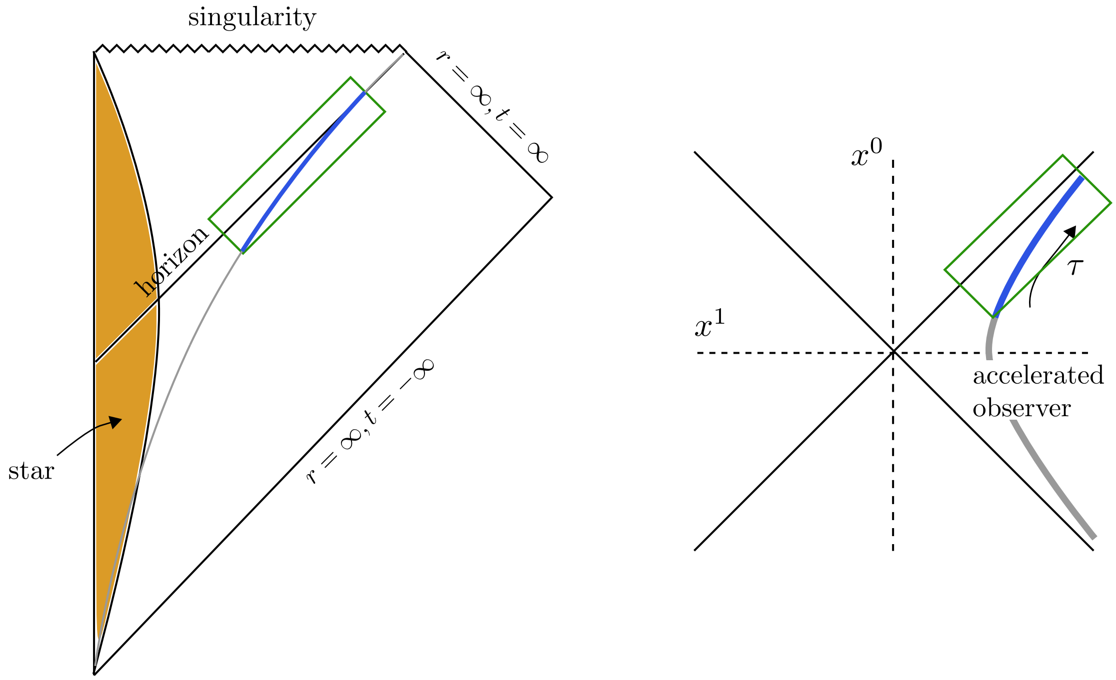


Figure 1: Left: Penrose diagram of a black hole formed by gravitational collapse. Right: Zoomed-in view of the flat near-horizon region, with the trajectory of a uniformly accelerated observer at $\rho = a^{-1}$.

the temperature that an accelerated observer feels is

$$T_{proper} = \frac{1}{2\pi\rho} = \frac{a}{2\pi} = \frac{\hbar}{k_B c}\frac{a}{2\pi} \tag{2.12}$$

where a is the proper acceleration and we also restored all the units in the last formula. Though this argument seems a bit formal, one can check that a physical accelerating thermometer would actually record this temperature [46].

Now, this is the proper temperature felt by an observer very close to the horizon. Notice that it is infinite at $\rho = 0$ and it decreases as we move away. This decrease in temperature is consistent with thermal equilibrium in the presence of a gravitational potential. In other words, for a spherically symmetric configuration, in thermal equilibrium, the temperature obeys the Tolman relation [48]

$$T_{proper}(r)\sqrt{-g_{\tau\tau}(r)} = \text{constant.} \tag{2.13}$$

This formula tracks the redshifting of photons as they climb a gravitational potential. It says that locations at a higher gravitational potential feel colder to a local observer. Using the polar-like coordinates (2.10) and (2.12) we indeed get a constant equal to $1/(2\pi)$. Since this formula is valid also in the full geometry (2.7), we can then use it to find the temperature that an observer far from the black hole would feel. We simply need to undo the rescaling of time we did just above (2.8) and go to large r where $g_{tt} = -1$ to find the temperature

$$T = T_{proper}(r \gg r_s) = \frac{1}{4\pi r_s} \ . \tag{2.14}$$

This is the Hawking temperature. It is the temperature measured by an observer that is more than a few Schwarzschild radii away from the black hole.

2.3 The Euclidean black hole

We will now expand a bit more on the connection between Euclidean time and thermodynamics. We will then use it to get another perspective on thermal aspects of black holes. Sometimes Euclidean time t_E is called imaginary time and Lorentzian time t is called real time because of the Wick rotation $t = it_E$ mentioned above.

There are different ways to see that imaginary-time periodicity is the same as a temperature. In a thermal state, the partition function is

$$Z = \mathrm{Tr}\,[e^{-\beta H}] \ . \tag{2.15}$$

Any observable such as $\mathrm{Tr}\,[\mathcal{O}(t)\mathcal{O}(0)e^{-\beta H}]$ is periodic under $t \to t + i\beta$, using $\mathcal{O}(t) = e^{iHt}\mathcal{O}e^{-iHt}$ and the cyclic property of the trace.

A more general argument in quantum field theory is to recast the trace as a path integral. Real-time evolution by e^{-iHt} corresponds to a path integral on a Lorentzian spacetime, so imaginary-time evolution, $e^{-\beta H}$, is computed by a path integral on a Euclidean geometry. The geometry is evolved for imaginary time β, and the trace tells us to put the same boundary conditions at both ends and sum over them. A path integral on a strip of size β with periodic boundary conditions at the ends is the same as a path integral on a cylinder. Therefore in quantum field theory $Z = \mathrm{Tr}\, e^{-\beta H}$ is calculated by a path integral on a Euclidean cylinder with $\theta = \theta + \beta$. Any observables that we calculate from this path integral will automatically be periodic in imaginary time.

Similarly, in a black hole spacetime, the partition function at inverse temperature β is calculated by a Euclidean path integral. The geometry is the Euclidean black hole, obtained from the Schwarzschild metric (2.7) by setting $t = it_E$,

$$ds_E^2 = \left(1 - \frac{r_s}{r}\right) dt_E^2 + \frac{dr^2}{1 - \frac{r_s}{r}} + r^2 d\Omega_2^2 \ , \qquad\qquad t_E = t_E + \beta \ . \tag{2.16}$$

In the Euclidean geometry, the radial coordinate is restricted to $r > r_s$, because we saw that $r - r_s$ is like the radial coordinate in polar coordinates, and $r = r_s$ is the origin − Euclidean

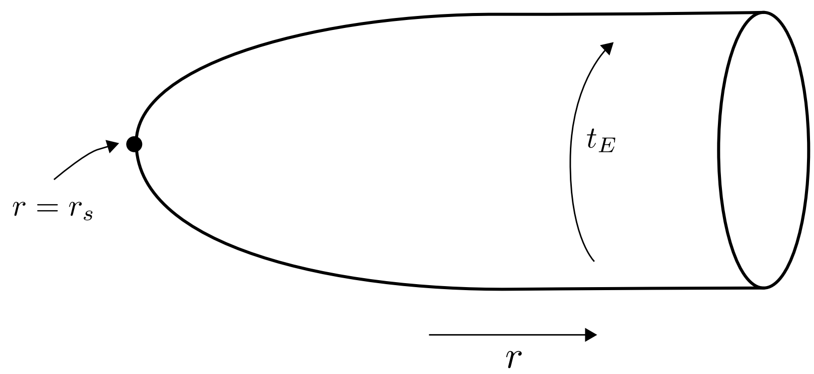


Figure 2: The Euclidean Schwarzschild black hole. The Euclidean time and radial directions have the geometry of a cigar, which is smooth at the tip, $r = r_s$. At each point we also have a sphere of radius r.

black holes do not have an interior. In order to avoid a conical singularity at $r = r_s$ we need to adjust β to

$$\beta = 4\pi r_s \ . \tag{2.17}$$

This geometry, sometimes called the 'cigar,' is pictured in fig. 2. The tip of the cigar is the horizon. Far away, for $r \gg r_s$, there is a Euclidean time circle of circumference β, which is the inverse temperature as seen by an observer far away. Notice that in the gravitational problem we fix the length of the circle far away, but we let the equations determine the right radius in the rest of the geometry.

The Euclidean path integral on this geometry is interpreted as the partition function,

$$Z(\beta) = \text{Path integral on the Euclidean black hole} \sim e^{-I_{\text{classical}}} Z_{\text{quantum}} \ . \tag{2.18}$$

It has contributions from both gravity and quantum fields. The gravitational part comes from the Einstein action, I, and is found by evaluating the action on the geometry (2.16). The quantum part is obtained by computing the partition function of the quantum fields on this geometry (2.16). It is important that the geometry is completely smooth at $r = r_s$ and therefore the quantum contribution has no singularity there. This is related to the fact that an observer falling into an evaporating black hole sees nothing special at the horizon, as in the classical theory.

Then applying the standard thermodynamic formula to the result,

$$S = (1 - \beta \partial_\beta) \log Z(\beta) \tag{2.19}$$

gives the generalized entropy (2.4). We will not give the derivation of this result but it uses that we are dealing with a solution of the equations of motion and that the non-trivial part of the variation can be concentrated near $r = r_s$ [49].

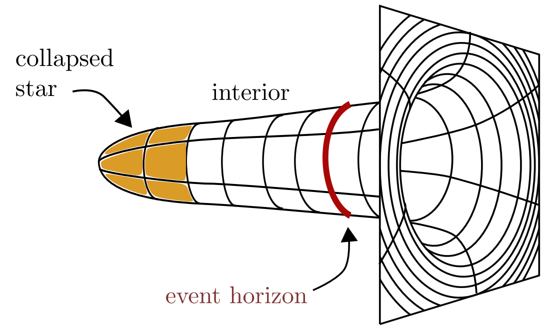
2.4 Evaporating black holes

Hawking radiation carries energy away to infinity and therefore reduces the mass of the black hole. Eventually the black hole evaporates away completely — a primordial black hole of mass 10^{12} kg, produced in the early universe, would evaporate around now. The Hawking temperature of a solar mass black hole is 10^{-7} K and its lifetime is 10^{64} years. The spacetime for this process is described in figures 3 and 4.

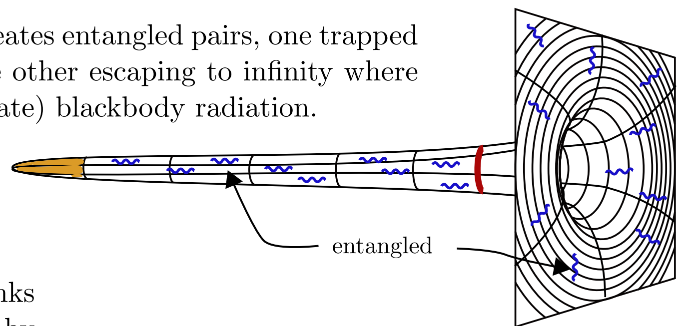
The Hawking process can be roughly interpreted as pair creation of entangled particles near the horizon, with one particle escaping to infinity and the other falling toward the singularity. This creation of entanglement is crucial to the story and we will discuss it in detail after introducing a few more concepts.

Stages of Black Hole Evaporation

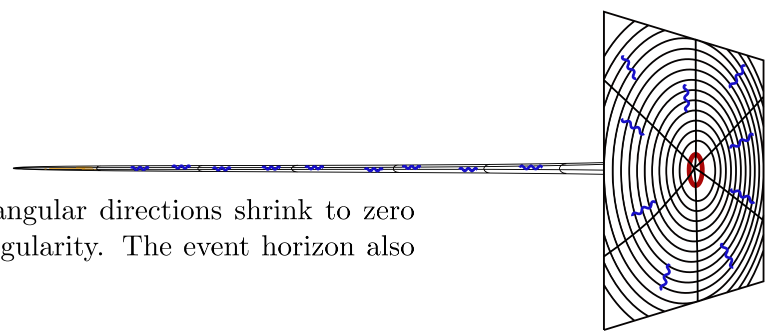
(a) After stellar collapse, the outside of the black hole is nearly stationary, but on the inside, the geometry continues to elongate in one direction while pinching toward zero size in the angular directions.



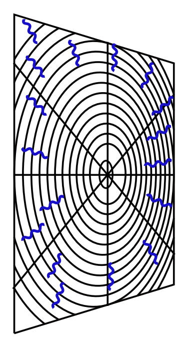
(b) The Hawking process creates entangled pairs, one trapped behind the horizon and the other escaping to infinity where it is observed as (approximate) blackbody radiation.



The black hole slowly shrinks as its mass is carried away by the radiation.



(c) Eventually the angular directions shrink to zero size. This is the singularity. The event horizon also shrinks to zero.



(d) At the end there is a smooth spacetime containing thermal Hawking radiation but no black hole.

Figure 3

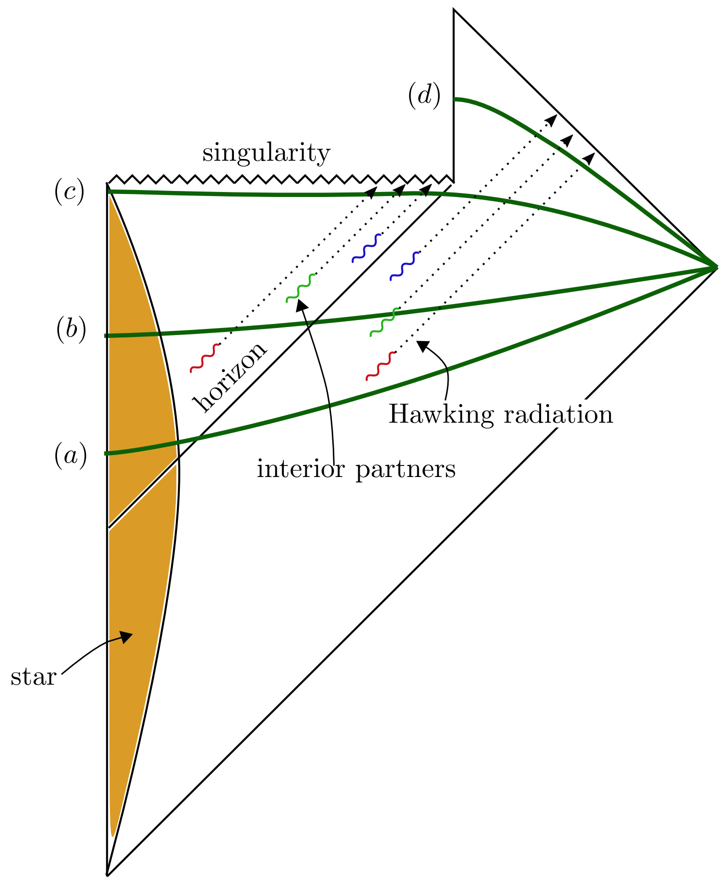


Figure 4: Penrose diagram for the formation and evaporation of a black hole. Spatial slices (a)-(d) correspond to the slices drawn in fig. 3.

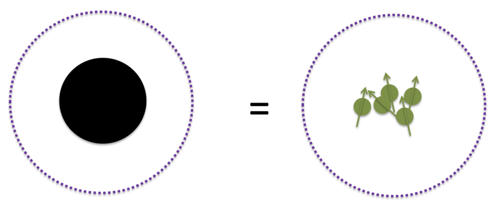
3 The black hole as an ordinary quantum system

The results we reviewed above suggest that the black hole can be regarded as an ordinary system, obeying the laws of thermodynamics. More precisely, as an object described by a finite, but large, number of degrees of freedom that obey the ordinary laws of physics, which in turn imply the laws of thermodynamics.

In fact, this has been such an important idea in the development of the subject that we will call it the "central dogma."

Central Dogma

> As seen from the outside, a black hole can be described in terms of a quantum system with Area$/(4G_N)$ degrees of freedom, which evolves unitarily under time evolution.



Using this hypothesis, the black hole and the whole spacetime around it, up to some surface denoted by the dotted circle, can be replaced by a quantum system. This quantum system interacts with the outside via a unitary Hamiltonian.

Let us make some remarks about this.

- Notice that it is a statement about the black hole as seen from the outside. There is no statement about the black hole interior, yet.
- The statement about the number of degrees of freedom is primarily a statement about the logarithm of the dimension of the Hilbert space. We make no distinction between qubits, fermions or other degrees of freedom. What is important is that the Hilbert space has *finite* dimension.
- The degrees of freedom that appear in this statement are not manifest in the gravity description. Some researchers have tried to see them as coming from the thermal atmosphere, by putting a cutoff, or "brick wall" at some Planck distance from the horizon [50, 51]. Unfortunately, such ideas remain vague since such cutoffs are not manifest from the gravity point of view, which treats the horizon as a smooth surface.

- Unitary evolution implies that we have a Hamiltonian that generates the time evolution. Again, this Hamiltonian is not manifest in the gravity description. The gravity description has a Hamiltonian constraint, which determines the bulk evolution. But this constraint is a property of the full spacetime, we do not know how to pull out a purely exterior part. In principle, the Hamiltonian could be very general. But the fact that it gives rise to the gravity evolution constrains some properties. For example, it should be strongly interacting and generate a very chaotic evolution.

- We said that the black hole evolves unitarily. This is when we surround the black hole by a reflecting wall and we consider the full system inside this wall. However, if the black hole lives in an asymptotically flat geometry, it is convenient to draw an imaginary surface surrounding the black hole and call everything inside the "quantum system" that appears in the central dogma. This quantum system is then coupled with the external degrees of freedom living outside this surface. We usually think of the region outside the imaginary surface as a quantum system in a fixed spacetime, where we ignore large fluctuations of the background, though we can still consider weakly interacting gravitons. The full coupled evolution should be unitary. In other words, in this context, the gravity answers are compared to those of a quantum system that is coupled to the degrees of freedom far from the black hole at this imaginary cutoff surface. Here we are imagining that this surface is at a few Schwarzschild radii from the black hole.

- Often people ask: how is a black hole different from a hot piece of coal? This central dogma is saying that, as long as you remain outside, it is not fundamentally different, in the sense that both are governed by a unitary Hamiltonian and have a finite number of degrees of freedom. Unlike a piece of coal, the black hole has an interior shrouded by an event horizon, and making it fully compatible with the exterior view is a non-trivial problem that has not been completely solved.

- The name "central dogma" was borrowed from biology where the central dogma talks about the information transfer from DNA and RNA to proteins. Here it is also a statement about information – quantum information. It involves a certain dose of belief, because it is not something we can derive directly from the gravity description. We can view it as an unproven assumption about the properties of a full theory of quantum gravity. It is also something that is not accepted by some researchers. In fact, Hawking famously objected to it.

- Notice that this statement is in stark contrast with a naive reading of the spacetime geometry. The spacetime geometry can be viewed as having two "asymptotic" regions. One is the obvious region outside, and the other is the future region near the singularity. See figure 5. Of course, the semiclassical gravity theory does not tell us how to evolve past this singularity, or even whether such evolution makes sense. From this point of view the interior is something we cannot access from the outside, but there is no obvious reason why some quantum information could not be lost here. In other words, if a black hole is a "hole in space" where things can get in and get lost, then the central dogma would be FALSE. In fact, this is one reason why some people think it is indeed false.

- Both the results on black hole thermodynamics, as well as the results on fine-grained entropies we will discuss later, are true properties of a theory of gravity coupled to quantum fields and do *not* require the validity of the central dogma. In other words, we are not assuming the central dogma in this review − we are providing evidence for it.

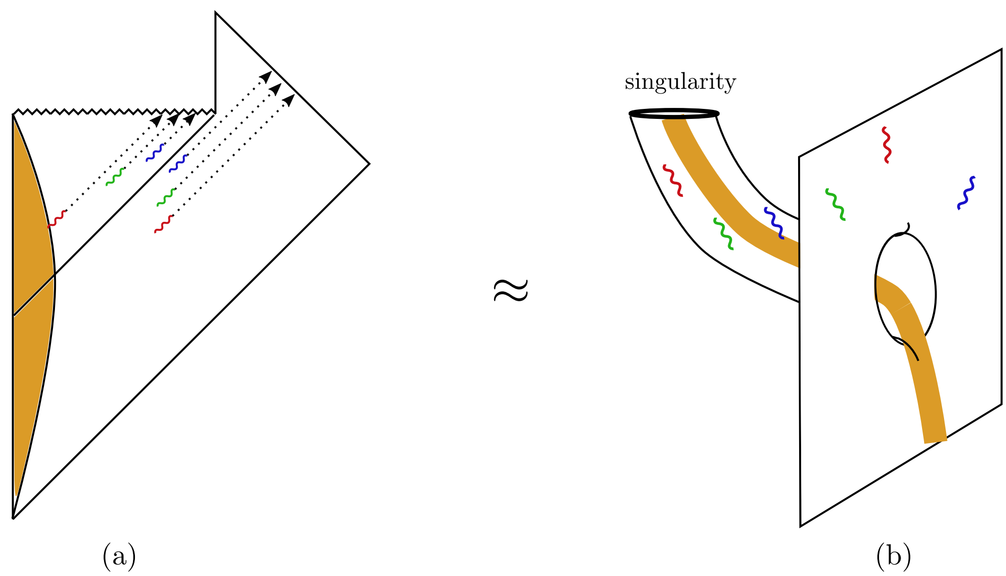


Figure 5: The skeptics' view: The diagram of an evaporating black hole is conceptually similar to one where we split off a baby universe, so that in the future we have two regions, the future region of the original universe, and the future of the interior, which is singular.

3.1 Evidence from string theory for the central dogma

Though we said that the "central dogma" is an unproven assumption, there is a great deal of very non-trivial evidence from string theory. String theory is a modification of Einstein gravity that leads to a well defined perturbative expansion and also some non-perturbative results. For this reason it is believed to define a full theory of quantum gravity.

One big piece of evidence was the computation of black hole entropy for special extremal black holes in supersymmetric string theories [52]. In these cases one can reproduce the Bekenstein-Hawking formula from an explicit count of microstates. These computations match not only the area formula, but all its corrections, see e.g. [53]. Another piece of evidence comes from the AdS/CFT correspondence [54–56], which is a conjectured relation between the physics of AdS and a dual theory living at its boundary. In this case, the black hole and its whole exterior can be represented in terms of degrees of freedom living at the boundary. There is also evidence from matrix models that compute scattering amplitudes in special vacua [57]. We will not discuss this further in this review, since we are aiming to explain features which rely purely on gravity as an effective field theory.

4 Fine-grained vs coarse-grained entropy

There are two notions of entropy that we ordinarily use in physics and it is useful to make sure that we do not confuse them in this discussion.

The simplest to define is the von Neuman entropy. Given the density matrix, ρ, describing the quantum state of the system, we have

$$S_{vN} = -Tr[\rho \log \rho] \tag{4.1}$$

It quantifies our ignorance about the precise quantum state of the system. It vanishes for a pure state, indicating complete knowledge of the quantum state. An important property is that it is invariant under unitary time evolution $\rho \to U\rho U^{-1}$.

The second notion of entropy is the coarse-grained entropy. Here we have some density matrix ρ describing the system, but we do not measure all observables, we only measure a subset of simple, or coarse-grained observables A_i. Then the coarse-grained entropy is given by the following procedure. We consider all possible density matrices $\tilde{\rho}$ which give the same result as our system for the observables that we are tracking, $Tr[\tilde{\rho}A_i] = Tr[\rho A_i]$. Then we compute the von Neumann entropy $S(\tilde{\rho})$. Finally we maximize this over all possible choices of $\tilde{\rho}$.

Though this definition looks complicated, a simple example is the ordinary entropy used in thermodynamics. In that case the A_i are often chosen to be a few observables, say the approximate energy and the volume. The thermodynamic entropy is obtained by maximizing the von Neumann entropy among all states with that approximate energy and volume.

Coarse-grained entropy obeys the second law of thermodynamics. Namely, it tends to increase under unitary time evolution.

Let us make some comments.

- The von Neumann entropy is sometimes called the "fine-grained entropy", contrasting it with the coarse-grained entropy defined above. Another common name is "quantum entropy."

- Note that the generalized entropy defined in (2.4) increases rapidly when the black hole first forms and the horizon grows from zero area to a larger area. Therefore if it has to be one of these two entropies, it can only be the thermodynamic entropy. In other words, the entropy (2.4) defined as the area of the horizon plus the entropy outside is the coarse-grained entropy of the black hole.

- Note that if we have a quantum system composed of two parts A and B, the full Hilbert space is $H = H_A \times H_B$. Then we can define the von Neumann entropy for the subsystem A. This is computed by first forming a density matrix ρ_A obtained by taking a partial trace over the system B. The entropy of ρ_A can be non-zero, even if the full system is in a pure state. This arises when the original pure state contains some entanglement between the subsystems A and B. In this case $S(A) = S(B)$ and $S(A \cup B) = 0$.

- The fine-grained entropy cannot be bigger than the coarse-grained entropy, $S_{vN} \leq S_{coarse}$. This is a simple consequence of the definitions, since we can always consider ρ as a candidate $\tilde{\rho}$. Another way to say this is that because S_{coarse} provides a measure of the total number of degrees of freedom available to the system, it sets an upper bound on how much the system can be entangled with something else.

It is useful to define the fine-grained entropy of the quantum fields in a region of space. Let Σ be a spatial region, defined on some fixed time slice. This region has an associated density matrix ρ_Σ, and the fine-grained entropy of the region is denoted

$$S_{vN}(\Sigma) \equiv S_{vN}(\rho_\Sigma) \ . \tag{4.2}$$

If Σ is not a full Cauchy slice, then we will have some divergences at its boundaries. These divergences are not important for our story, they have simple properties and we can deal with them appropriately. Also, when Σ is a portion of the full slice, $S_{vN}(\Sigma)$ is generally time-dependent. It can increase or decrease with time as we move the slice forwards in time. The slice Σ defines an associated causal diamond, which is the region that we can determine if we know initial data in Σ, but not outside Σ. The entropy is the same for any other slice $\tilde{\Sigma}$ which has the same causal diamond as Σ, see figure 6.

4.1 Semiclassical entropy

We now consider a gravity theory which we are treating in the semiclassical approximation. Namely, we have a classical geometry and quantum fields defined on that classical geometry.

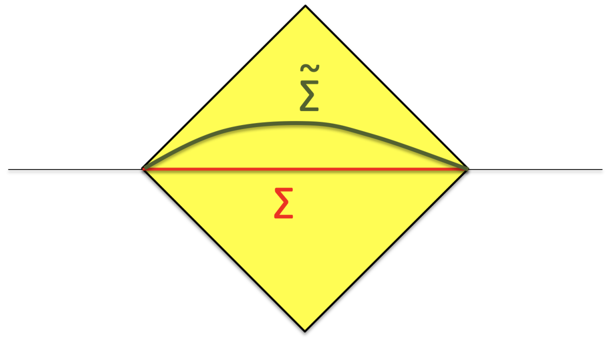


Figure 6: Given a region Σ of a spatial slice, shown in red, we can define its causal diamond to be all points where the evolution is uniquely determined by initial conditions on Σ. The alternative slice $\tilde{\Sigma}$ defines the same causal diamond. The von Neumann entropies are also the same.

Associated to a spatial subregion we can define its "semiclassical entropy," denoted by

$$S_{\text{semi-cl}}(\Sigma) \ . \tag{4.3}$$

$S_{\text{semi-cl}}$ is the von Neumann entropy of quantum fields (including gravitons) as they appear on the semiclassical geometry. In other words, this is the fine-grained entropy of the density matrix calculated by the standard methods of quantum field theory in curved spacetime. In the literature, this is often simply called the von Neumann entropy (it is also called S_{matter} or S_{outside} in the black hole context).

5 The Hawking information paradox

The Hawking information paradox is an argument against the "central dogma" enunciated above [58]. It is only a paradox if we think that the central dogma is true. Otherwise, perhaps it can be viewed as a feature of quantum gravity.

The basic point rests on an understanding of the origin of Hawking radiation. We can first start with the following question. Imagine that we make a black hole from the collapse of a pure state, such as a large amplitude gravity wave [59]. This black hole emits thermal radiation. Why do we have these thermal aspects if we started with a pure state? The thermal aspects of Hawking radiation arise because we are essentially splitting the original vacuum state into two parts, the part that ends up in the black hole interior and the part that ends up in the exterior. The vacuum in quantum field theory is an entangled state. As a whole state it is pure, but the degrees of freedom are entangled at short distances. This implies that if we only consider half of the space, for example half of flat space, we will get a mixed state on that half. This is a very basic consequence of unitarity and relativistic invariance [47]. Often this is explained qualitatively as follows. The vacuum contains pairs of particles that are constantly being created and annihilated. In the presence of a horizon, one of the members of the pair can go to infinity and the other member is trapped in the black hole interior. We will call them the "outgoing Hawking quantum" and the "interior Hawking quantum." These two particles are entangled with each other, forming a pure state. However if we consider only one member, say the outgoing Hawking quantum, we fill find it in a mixed state, looking like a thermal state at the Hawking temperature (2.14). See figure 3b and figure 4.

This process on its own does not obviously violate the central dogma. In fact, if we had a very complex quantum system which starts in a pure state, it will appear to thermalize and will emit radiation that is very close to thermal. In particular, in the early stages, if we computed the von Neumann entropy of the emitted radiation it would be almost exactly thermal because the radiation is entangled with the quantum system. So it is reasonable to expect that during the initial stages of the evaporation, the entropy of radiation rises. However, as the black hole evaporates more and more, its area will shrink and we run into trouble when the entropy of radiation is bigger than the thermodynamic entropy of the black hole. The reason is that now it is not possible for the entropy of radiation to be entangled with the quantum system describing the black hole because the number of degrees of freedom of the black hole is given by its thermodynamic entropy, the area of the horizon. In other words, if the black hole degrees of freedom together with the radiation are producing a pure state, then the fine-grained entropy of the black hole should be equal to that of the radiation $S_{\text{black hole}} = S_{\text{rad}}$. But this fine-grained entropy of the black hole should be less than the Bekenstein-Hawking or thermodynamic entropy of the black hole, $S_{\text{black hole}} \leq S_{\text{Bekenstein}-\text{Hawking}} = S_{\text{coarse}-\text{grained}}$.

If the central hypothesis were true, we would expect that the entropy of radiation would need to start decreasing at this point. In particular, it can never be bigger than the Bekenstein Hawking entropy of the old black hole. Notice that we are talking about the von Neumann or fine-grained entropy of the radiation. Then, as suggested by D. Page [60, 61], the entropy of the radiation would need to follow the curve indicated in figure 7, as opposed to the Hawking curve. The time at which $S_{\text{Bekestein}-\text{Hawking}} = S_{\text{rad}}$ is called the Page time.

Now let us finish this discussion with a few comments.

- Note that, as the black hole evaporates, its mass decreases. This is sometimes called the "backreaction" of Hawking radiation. This effect is included in the discussion. And it does not "solve" the problem.
- When the black hole reaches the final stages of evaporation, its size becomes comparable to the Planck length and we can no longer trust the semiclassical gravity description.

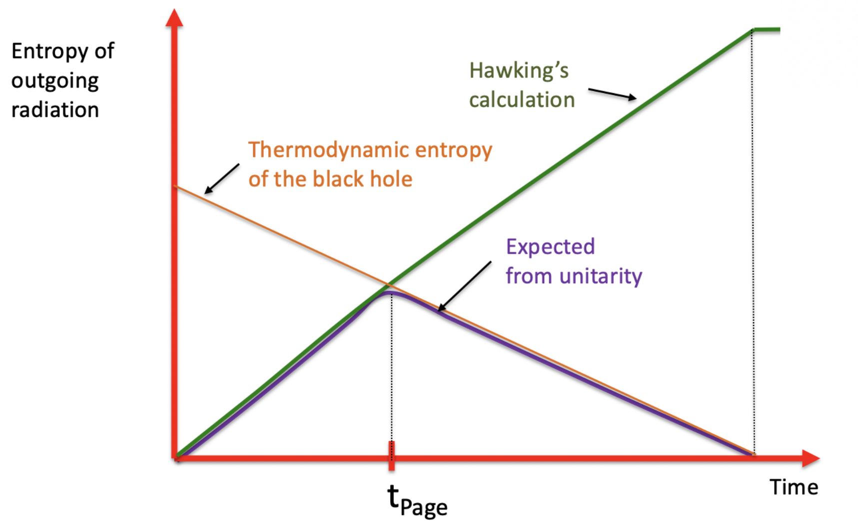


Figure 7: Schematic behavior of the entropy of the the outgoing radiation. The precise shape of the lines depends on the black hole and the details of the matter fields being radiated. In green we see Hawking's result, the entropy monotonically increases until t_{End}, when the black hole completely evaporates. In orange we see the thermodynamic entropy of the black hole. If the process is unitary, we expect the entropy of radiation to be smaller than the thermodynamic entropy. If it saturates this maximum, then it should follow the so called "Page" curve, denoted in purple. This changes relative to the Hawking answer at the Page time, t_{Page}, when the entropy of Hawking radiation is equal to the thermodynamic entropy of the black hole.

This is not relevant since the conflict with the central dogma appeared at the Page time, when the black hole was still very big.

- The argument is very robust since it relies only on basic properties of the fine-grained entropy. In particular, it is impossible to fix the problem by adding small corrections to the Hawking process by slightly modifying the Hamiltonian or state of the quantum fields near the horizon [62–64]. In other words, the paradox holds to all orders in perturbation theory, and so if there is a resolution it should be non-perturbative in the gravitational coupling G_N.

- We could formulate the paradox by constantly feeding the black hole with a pure quantum state so that we exactly compensate the energy lost by Hawking radiation. Then the mass of the black hole is constant. Then the paradox would arise when this process goes on for a sufficiently long time that the entropy of radiation becomes larger than the entropy of the black hole.

- One could say that the gravity computation only gives us an approximate description

> and we should not expect that a sensitive quantity like the von Neumann entropy should be exactly given by the semiclassical theory. In fact, this is what was said until recently. We will see however, that there *is* a way to compute the von Neuman entropy using just this semiclassical description.

We have described here one aspect of the Hawking information paradox, which is the aspect that we will see how to resolve. We will comment about other aspects in the discussion.

6 A formula for fine-grained entropy in gravitational systems

As we mentioned above, the Bekenstein Hawking entropy formula should be viewed as the coarse-grained entropy formula for the black hole, since it increases under time evolution. This is clear when the black hole first forms and has not yet had time to emit Hawking radiation.

Surprisingly there is also a gravitational formula for the von Neumann or fine-grained entropy [5,6,9,10]. It is also given by a formula that involves a generalized entropy, with an area plus the entropy of fields outside. The only difference is in the choice of the dividing surface. Roughly the idea is that we choose a surface such that the generalized entropy is minimized. This minimal value is the fine-grained entropy,

$$S \sim \min \left[\frac{\text{Area}}{4G_N} + S_{\text{outside}} \right] \ . \tag{6.1}$$

Now, (6.1) captures the spirit of the formula, but the precise formula is slightly more complicated. The reason is the following. A surface is a codimension-2 object. This means that it has two dimensions less than that of the full spacetime. In our case it is localized along one of the spatial dimensions and also in time. We are looking for a surface that minimizes (6.1) in the spatial direction but maximizes it in the time direction. So we really should look for "extremal surfaces" by moving them both in space and in time. If there are many extremal surfaces we should find the global minimum. Another equivalent definition is the following maxi-min construction [65,66]. First choose a spatial slice (a Cauchy slice) and find the minimal surface. Then find the maximum among all choices of the Cauchy slice. Then a more precise version of the formula is [5,6,9,10]

$$\boxed{S = \min_X \left\{ \text{ext}_X \left[\frac{\text{Area}(X)}{4G_N} + S_{\text{semi-cl}}(\Sigma_X) \right] \right\}} \ , \tag{6.2}$$

where X is a codimension-2 surface, Σ_X is the region bounded by X and the cutoff surface, and $S_{\text{semi-cl}}(\Sigma_X)$ is the von Neumann entropy of the quantum fields on Σ_X appearing in the