

Introduction to black hole thermodynamics

Pietro Benetti Genolini

*Département de Physique Théorique, Université de Genève,
24 quai Ernest-Ansermet, 1211 Genève, Suisse*

E-mail: pietro.benettigenolini@unige.ch

ABSTRACT: These are the lecture notes for a course in the *Roberto Salmeron School in Mathematical Physics* held at the University of Brasilia in September 2025. The course provides a concise and biased introduction to black hole thermodynamics. It covers the laws of black hole mechanics, Unruh and Hawking effects, gravitational path integral, and AdS black holes.

Updated on September 8, 2025

Contents

Introduction	2
1 The laws of black hole mechanics	5
1.1 Rindler horizon	5
1.2 Schwarzschild horizon	9
1.3 Killing horizons	13
2 Hawking radiation and black hole thermodynamics	17
2.1 The laws of black hole thermodynamics	17
2.2 Quantum field theory on curved spaces	19
2.3 Black holes	27
2.4 Near the horizon of a black hole	28
2.5 Where to now?	33
3 The gravitational path integral	36
3.1 Definition	36
3.2 Gravity action	38
3.3 Hawking–Page transition	44
4 Selected topics on the gravitational path integral	53
4.1 Entropy from topology	53
4.2 Rotation, charge and complex metric	53
4.3 Subtleties	58

Introduction

General relativity is a very successful theory, but it is also a fundamentally classical theory, in the sense that no probabilities are involved in discussing observables. A simple dimensional argument tells us that its description of the physics breaks down at the scale of the Planck energy, which is the unique combination of the fundamental constants involved in the game, $E_P \equiv \sqrt{\hbar c^5/G} \sim 10^{19}$ GeV. Its extraordinary size is due to the weakness of gravity. For reference, the typical energy scale probed by the LHC is $E_W \sim 10^4$ GeV. Therefore, we definitely shall not see quantum gravity effects in everyday life or any experiment currently built on Earth. Why should we care, then?

General relativity itself tells us to do so. Spacetime singularities are necessarily present in classical solutions describing either gravitational collapse or cosmology. At these singularities, the classical theory is incomplete, as it doesn't predict a way of prescribing boundary conditions. Therefore, a theory of quantum gravity is needed.

Black holes provide an arena where quantum gravity effects come sharply into focus. They are a striking prediction of general relativity: regions of spacetime bounded by a horizon out of which nothing can escape, and inside which lies hidden a spacetime singularity. Yet, they are not just exotic curiosities, rather the result of quite generic gravitational collapse, and they are real astrophysical objects observed in our universe.

Classically, a black hole is the perfect absorber, as it can emit nothing. However, its defining features (geometry of the horizon, conserved charges) are related by a set of equations that bear a striking resemblance to the laws of thermodynamics.

The analogy becomes physical once one includes quantum effects. Hawking studied a black hole surrounded by a quantum field, and famously showed that it behaves like a thermal object, whose temperature and entropy are fixed by the geometry of the horizon.

Since then, understanding the thermodynamics of black holes has been one of the key questions guiding research into quantum gravity, and it has led, among others, to profound insights into quantum field theory in curved spacetime, the discovery of the holographic nature of gravity, the development of gravitational path integral techniques, and the statistical interpretation of black hole entropy.

These notes provide a very concise and very biased introduction to the subject of black hole thermodynamics.

We begin in section 1 with a review of some results in the classical description of black holes, focusing in particular on the properties of Killing horizons. In section 2, we use techniques from thermal quantum field theory to argue that an observer along the orbit of a Killing vector with a Killing horizon would detect a temperature that is directly related to the geometry of the horizon itself. We cover the Unruh effect, originally concerning accelerated observers in flat space, and summarize the Hawking effect, which is instead a phenomenon related to more general event horizons arising from gravitational collapse. In section 3, we introduce a framework for the quantization of gravity, the gravitational path integral, and we look at the thermodynamics of gravity in anti-de Sitter spacetime. Finally, in section 4, we conclude with a brief review of additional topics related to the gravitational path integral, and issues that had been swept under the rug in section 3.

The literature on the topics of the course is enormous, a large number of lecture notes and books exist, and I have drawn from them. I make no claim of originality for the contents of the notes, besides the fact that the choice of topics and the presentation is biased by my own (perhaps idiosyncratic) tastes. Most of the topics were developed within the span of a decade between mid 1970s and mid 1980s. Many of the original papers are beautifully written, are still relevant today and well worth reading. I try to refer to the most relevant ones as we go along, and some have been reprinted in a single book in [GH93]. For the global structure of the notes, and the presentation, I was very influenced by the lecture notes by Harvey Reall [Rea20] and Simon Ross [Ros05].

As for additional lecture notes and books, I have found the following resources useful.

On black hole mechanics and thermodynamics via quantum field theory on curved spacetime (sections 1 and 2)

- Birrell, Davies, *Quantum Fields in Curved Space* [BD84]
- Fulling, Ruijsenaars, *Temperature, periodicity and horizons* [FR87]
- Jacobson, *Introduction to quantum fields in curved space-time and the Hawking effect* [Jac03]

- Reall, *Part 3 Black Holes* [[Rea20](#)]
- Ross, *Black hole thermodynamics* [[Ros05](#)]
- Townsend, *Black holes: Lecture notes* [[Tow97](#)]
- Wald, *Quantum Field Theory in Curved Space-Time and Black Hole Thermodynamics* [[Wal95](#)]
- Witten, *Introduction to black hole thermodynamics* [[Wit24](#)]

On Euclidean quantum gravity approach and applications (sections [3](#) and [4](#))

- Hawking, *Euclidean Quantum Gravity* [[Haw78](#)]
- Hawking, *The path integral approach to quantum gravity* [[Haw79](#)]

1 The laws of black hole mechanics

1.1 Rindler horizon

We begin, counterintuitively for a course on black holes, with flat two-dimensional space

$$ds^2 = -dt^2 + dx^2, \quad (1.1)$$

and we look at an observer experiencing a constant acceleration α (in her frame). Her worldline parametrized by proper time s is (with an appropriate choice of origin)

$$t = \frac{1}{\alpha} \sinh \alpha s, \quad x = \frac{1}{\alpha} \cosh \alpha s, \quad -t^2 + x^2 = \alpha^{-2}, \quad (1.2)$$

so it's the hyperbola with asymptotes $\{t = x, t = -x\}$ represented in Figure 1. The tangent to the worldline is

$$b = (\sinh \alpha s, \cosh \alpha s) = \alpha \left(x \frac{\partial}{\partial t} + t \frac{\partial}{\partial x} \right), \quad (1.3)$$

with length $b^2 = -1$, and the magnitude of the proper acceleration $A_a = \nabla_b b_a$ is indeed α^2 . From our frame, we see her approach (and never get to) the speed of light, and in her frame, eventually she's going to be able to receive information from the entire region $\{x > t\}$, but nothing left of the line $\{x = t\}$, which is why we refer to this line as the *Rindler horizon*. This shows that the physics measured by the accelerated observer is quite different from that seen by an inertial one, e.g. with worldline $\{t = s, x = 0\}$, but there is a easy way of getting rid of the Rindler horizon: she could simply stop accelerating.

Now we generalize slightly, and consider now the family of observers corresponding to all the orbits of b^a in (1.3). These are the hyperbolas $\{-t^2 + x^2 = \text{constant}\}$, including the degenerate case of the straight asymptotes $\{t = x\}$, $\{t = -x\}$. We remark two important properties of b^a : first, since we are in flat space, it is easy to notice that

$$\nabla_\mu b_\nu = \alpha \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad (1.4)$$

so b^a is a Killing vector (in fact, it's the generator of boosts). Moreover, its length is not constant: $b^2 = -\alpha^2(x^2 - t^2)$. This signals that the observers are not “free.” It's a

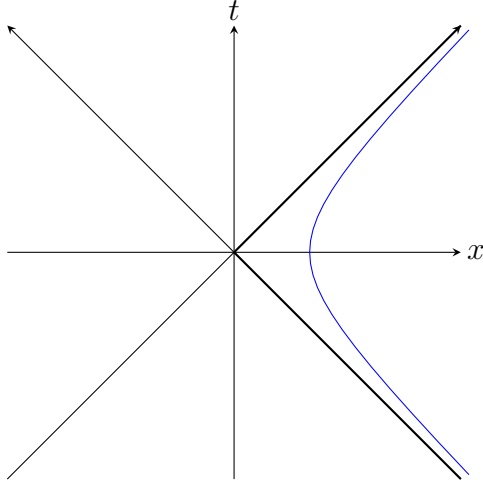


Figure 1: The worldline of an accelerated observer in flat space.

general property of motion along a Killing vector field ξ^a that it describes a geodesic if and only if the norm is constant, since

$$\xi^b \nabla_b \xi^a = -\xi^b \nabla_a \xi_b = -\frac{1}{2} \nabla_a \xi^2, \quad (1.5)$$

so the integral curve of ξ^a is an affinely parametrized geodesic if and only if ξ^2 is constant. In the case of the generator of the boosts, it's timelike only on the two wedges R and L in Figure 2, null on the two lines $\mathcal{N} \equiv \{t = x\} \cup \{t = -x\}$, with the origin being a special point where b^a vanishes, and becomes spacelike in the top and bottom wedges (this is also clear from the orbits).

Focus on the wedges R and L, where b^a is timelike, and consider an observer along an orbit of b^a , with normalized velocity $u^a = \frac{1}{\sqrt{-b^2}} b^a$. Her proper acceleration is

$$A_a = \nabla_u u_a = \frac{1}{\sqrt{-b^2}} b^b \nabla_b \frac{b_a}{\sqrt{-b^2}} = \frac{1}{-b^2} b^b \nabla_b b_a = -\frac{1}{-b^2} b^b \nabla_a b_b = \nabla_a \log \sqrt{-b^2}, \quad (1.6)$$

with magnitude

$$A = \frac{1}{\sqrt{x^2 - t^2}}.$$

So, the magnitude of the proper acceleration measured by each observer is constant, though it changes from orbit to orbit. In particular, it's α^2 on the orbit we started with, where $b^2 = -1$, vanishes at spatial infinity ($x \rightarrow \infty$), and diverges on the asymptotes \mathcal{N} . This is the acceleration measured by an accelerometer carried by each observer,

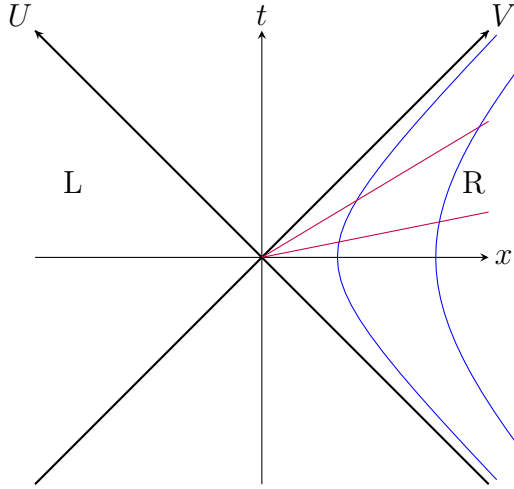


Figure 2: Minkowski spacetime with the wedges determined by the norm of the boost Killing vector b^a . In blue are two loci of constant ξ , which are trajectories of the Rindler observers following orbits of b^a . In purple are two loci of constant η .

but it’s not the acceleration measured by an observer “at infinity.” We introduce this in a more general way that will be useful later.

If k^a is a timelike Killing vector field (in an asymptotically flat spacetime), and there’s a particle with velocity u , we define the “energy per unit mass measured at infinity” to be $E_\infty = -u \cdot k$. It is a constant if the particle moves along a geodesic, but we’re interested in the motion along an accelerated trajectory. Interestingly, we are in fact considering motion along an orbit of k^a itself (i.e., a stationary observer), so $E_\infty = \sqrt{-k^2}$, in which case

$$\nabla_a E_\infty = \nabla_a \sqrt{-k^2} = \sqrt{-k^2} A_a, \quad (1.7)$$

where A_a is the proper (local) acceleration measured by the stationary observer, as showed in (1.6). By construction, this is the force per unit mass measured at infinity, so we find the relation

$$A_\infty = \sqrt{-k^2} A. \quad (1.8)$$

We can interpret this relation physically: the acceleration of the stationary observer requires a force, which we can imagine being provided by some observer at infinity. The force measured locally by the accelerating observer will be different from that of the observer at infinity, because of a “redshift” factor due to the gravitational pull. In

the case of the Rindler observer, we find that

$$A_\infty = \alpha. \quad (1.9)$$

Importantly, this is constant even as $x \rightarrow |t|$, we get closer to the horizon, and the local acceleration diverges.

Another important property of b^a is that it is normal to \mathcal{N} . This can be more easily seen introducing the light-cone coordinates

$$U = t - x, \quad V = t + x, \quad \Rightarrow \quad ds^2 = -dUdV, \quad b = \alpha \left(V \frac{\partial}{\partial V} - U \frac{\partial}{\partial U} \right). \quad (1.10)$$

Indeed

$$b_a = \frac{\alpha}{2} (-V dU + U dV)_a = \begin{cases} -\frac{\alpha}{2} V (dU)_a & U = 0 \\ \frac{\alpha}{2} U (dV)_a & V = 0 \end{cases}, \quad (1.11)$$

which shows that it's normal to \mathcal{N} .¹

Since $b^2|_{\mathcal{N}} = 0$, its derivative is normal to \mathcal{N} and hence proportional to b_a : indeed we have

$$b^2 = \alpha^2 VU \quad \Rightarrow \quad \nabla_a b^2 = \begin{cases} -2\alpha b_a & U = 0 \\ 2\alpha b_a & V = 0 \end{cases}. \quad (1.12)$$

Finally, before leaving our accelerating observer, we define a set of adapted coordinates, that is, coordinates (η, ξ) such that $b = \partial_\eta$: one such choice that covers the wedge R is

$$t = \frac{e^{\alpha\xi}}{\alpha} \sinh \alpha\eta, \quad x = \frac{e^{\alpha\xi}}{\alpha} \cosh \alpha\eta, \quad \eta = \frac{1}{\alpha} \log \frac{x+t}{x-t}, \quad \xi = \frac{1}{\alpha} \log \alpha \sqrt{x^2 - t^2}. \quad (1.13)$$

Note that η and ξ range from $-\infty$ to $+\infty$, though they only cover the wedge R: going back to figure 2, the hyperbolas in blue are the orbits of $b = \partial_\eta$ corresponding to constant ξ , and the straight purple lines correspond to constant η . In these coordinates, the Minkowski metric (1.1) has the form

$$ds^2 = e^{2\alpha\xi} (-d\eta^2 + d\xi^2). \quad (1.14)$$

Looking at this form, we confirm indeed that these coordinates, though defined along the worldline of the observer, are not inertial, as we would expect, since the observer

¹If f is a function, the normal to the set $\{f = \text{constant}\}$ is proportional to df .

is accelerated. However, as it's always the case, they are locally inertial: here, in a neighbourhood of $\xi = 0$, corresponding to the original worldline (1.2), $x^2 - t^2 = \alpha^{-2}$.

Rindler coordinates are often presented in a slightly different way, trading ξ for $\rho = e^{\alpha\xi}/\alpha > 0$. Then, the metric has the form

$$ds^2 = -\alpha^2 \rho^2 d\eta^2 + d\rho^2. \quad (1.15)$$

In these coordinates, which cover the right wedge, the Killing horizon is at $\rho = 0$. We shall see (1.15) appear when looking close to the event horizon of a black hole in section 2.4.

1.2 Schwarzschild horizon

The geometry of the Rindler horizon is characteristic of black holes. Look at the Schwarzschild metric²

$$ds^2 = -f(r)dt^2 + \frac{dr^2}{f(r)} + r^2 d\Omega_2^2, \quad (1.16)$$

$$f(r) \equiv 1 - \frac{2M}{r}, \quad ds^2(S^2) = d\theta^2 + \sin^2 \theta d\phi^2,$$

where the coordinates have ranges $t \in \mathbb{R}$, $r > 2M$, $\theta \in [0, \pi)$, $\phi \in [0, 2\pi)$ (the latter two covering a 2-sphere). There is a timelike Killing vector field $k = \partial_t$, with non-constant norm $k^2 = -f(r)$. This means that the motion of the stationary observer along k^a is not “free,” as it is not along geodesics.

Famously, the apparent singularity at $r = 2MG$ is only a coordinate singularity, and one can construct a maximal extension covered by the Kruskal–Szekeres coordinates (U, V, θ, ϕ) with³

$$ds^2 = -\frac{32M^3 e^{-r(U,V)/(2M)}}{r(U,V)} dU dV + r(U,V)^2 d\Omega_2^2, \quad (1.17)$$

where $U, V \in \mathbb{R}$ and $r(U, V)$ is the unique solution to

$$UV = -e^{r/(2M)} \left(\frac{r}{2M} - 1 \right). \quad (1.18)$$

²In fact, the Schwarzschild metric is the most general spherically symmetric solution to the Einstein equations (a statement going under the name of Birkhoff theorem), so it doesn't just describe a spherically symmetric static black hole, but also the outside of any spherically symmetric configuration, including a spherical gravitational collapse.

³Coordinates covering the maximal analytical extension had also been independently introduced by Synge (1950) and Fronsdal (1959), before Kruskal (1960) and Szekeres (1960) [MTW73].

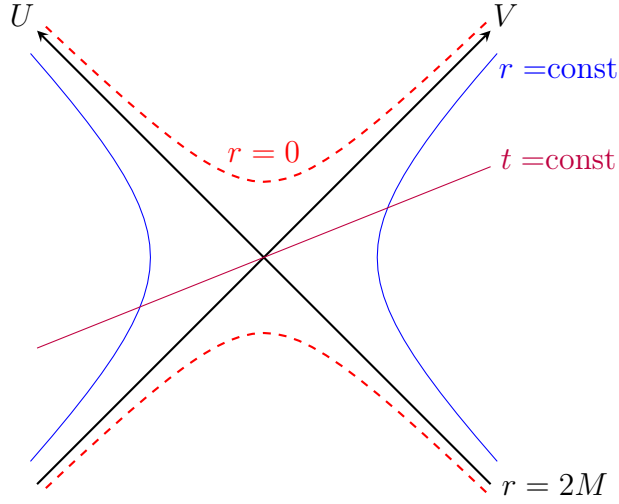


Figure 3: Maximal analytic extension of the Schwarzschild black hole, covered by the coordinates (U, V) (each point represents a two-sphere). In blue is a locus of constant $r > 2M$, corresponding to an orbit of $k = \partial_t$. In purple is a locus of constant t . The event horizon $\{r = 2M\}$ is the union of the two axes $\{U = 0\} \cup \{V = 0\}$, and the singularity $r = 0$ is in red.

A diagram of the $U - V$ plane is in Figure 3, where each point represents a two-sphere in the four-dimensional geometry, and the subset covered by the original Schwarzschild coordinates in (1.16) is the right wedge $\{U < 0, V > 0\}$. We have also drawn an orbit of k^a at constant $r > 2M$ (in blue) and a straight line corresponding to constant t . Looking at (1.18), we see that the red hyperbola $UV = 1$ corresponds to $r = 0$, whereas the axes $UV = 0$ correspond to $\{r = 2M\}$, the coordinate singularity in (1.16). Note the similarity between the worldlines of the stationary observers along k in Figure 3 and those along b in Figure 2. Again, there is a part of spacetime (namely $U > 0$) from which no information can reach an observer along on orbit of k with $U < 0$, so $\{U = 0\}$ acts as a horizon.

It's also straightforward to compute the proper acceleration and the acceleration measured at infinity using (1.6) and (1.8)

$$A = \frac{1}{\sqrt{f(r)}} \frac{M}{r^2}, \quad A_\infty = \frac{M}{r^2}. \quad (1.19)$$

As for the Rindler horizon, A diverges near the Schwarzschild horizon, meaning that the local acceleration measured by an observer near the Schwarzschild horizon would

diverge. However, if this observer was held by an observer “at infinity,” the latter one would only measure a finite value $A_\infty = \frac{1}{4M}$.

To conclude the analogies between the two spacetime, we remark that in the Kruskal coordinates, k^a has the form

$$k = \frac{1}{4M} \left(V \frac{\partial}{\partial V} - U \frac{\partial}{\partial U} \right), \quad (1.20)$$

which is entirely analogous to (1.10), including the fact that the constant prefactor is $A_\infty|_{\text{horizon}}$. We see that k^a is a well-defined Killing vector field on the entire Kruskal spacetime, its norm is $k^2 = f(r)$ even on the extension, and so it’s timelike for $r > 2M$, spacelike if $0 < r < 2M$, and null on $\mathcal{N} = \{r = 2M\} = \{UV = 0\}$, with a special locus at $\{U = V = 0\}$ where the Killing vector vanishes *tout court*. It is then straightforward to find the analog of (1.11), showing that k^a is again normal to \mathcal{N} , and to show that

$$\nabla_a k^2 = \begin{cases} -\frac{1}{2M} k_a & U = 0 \\ \frac{1}{2M} k_a & V = 0 \end{cases}. \quad (1.21)$$

Therefore, we have found an analogous structure in the two cases, which is worth defining more generally. We define a *Killing horizon* to be a null hypersurface \mathcal{N} such that there is a Killing vector ξ^a normal to it. In fact, we are mostly interested in a special class of Killing horizons: a *bifurcate Killing horizon* is the union of two null hypersurfaces that are both Killing horizons intersecting at a codimension-2 spacelike surface, the *bifurcation surface*, where the Killing vector ξ vanishes.

Since $\xi^2 = 0$ on a Killing horizon, its gradient must be proportional to ξ_a itself (which is normal to \mathcal{N}): we refer to the proportionality constant κ as *surface gravity*

$$\nabla_a \xi^2|_{\mathcal{N}} = -2\kappa \xi_a|_{\mathcal{N}} \quad \Rightarrow \quad \xi^b \nabla_b \xi^a|_{\mathcal{N}} = \kappa \xi^a|_{\mathcal{N}}. \quad (1.22)$$

The name is justified by the fact that κ is the limit at the horizon of the force per unit mass measured at infinity (1.8).

Notice that the surface gravity is not a property of the Killing horizon alone, as it depends on the normalization of the Killing vector: \mathcal{N} is a Killing horizon also for $c\xi^a$ for any real c , and the surface gravity would be $c\kappa$. Therefore, in order to assign a physical meaning to the surface gravity, we should also specify a normalization for ξ^a .

In the two cases considered earlier, we find from (1.12) and (1.21) that⁴

$$\begin{array}{ll} \text{Rindler horizon} & b^2 = -\alpha^2(x^2 - t^2), \quad \kappa = \pm\alpha, \\ \text{Schwarzschild horizon} & k^2 = -1 + \frac{2M}{r}, \quad \kappa = \pm\frac{1}{4M}. \end{array} \quad (1.23)$$

Note the difference: in the Rindler case, there is no canonical normalization of the Killing vector defining the horizon, which is there in the Schwarzschild case, since k^a generates (future-directed) time translations, normalized to have length -1 “at infinity.”

So, we see that though both cases have the structure of a bifurcate Killing horizon, they are different. This is good: one spacetime is flat and the other isn’t, so shouldn’t describe the same physics. How can we formalize this intuition?

First, consider a family of observers, i.e. a family of (inextendible) timelike curves $\{\gamma_\alpha\}$, then a non-empty boundary of the chronological pasts of their union $\partial I^-(\bigcup_\alpha \gamma_\alpha)$ is called a *future event horizon*. The future event horizon of the family of observers along orbits of b^a in the right wedge of Minkowski space in Figure 2 is $\{U = 0, V > 0\}$. The same equation also describes the future event horizon of the family of observers along orbits of k^a as in Figure 3. However, physically, we know that there is a big difference between the two cases: in Minkowski space, if the observer stops accelerating and moves on a geodesic, e.g. an orbit of ∂_t , then her chronological past includes the entire spacetime. In contrast, in the Kruskal spacetime there is no observer that escapes at arbitrarily large distances at arbitrarily late times that is able to receive information from the region $\{0 < r < 2M\}$. Slightly more formally, we introduce a notion of asymptotic (null) infinity \mathscr{I}^+ , where null geodesics end, and define a black hole as the region of spacetime that doesn’t belong to the chronological past of \mathscr{I}^+ , that is

$$\mathcal{B} \equiv M \setminus I^-(\mathscr{I}^+). \quad (1.24)$$

The future event horizon is the boundary $\partial\mathcal{B}$. In the Kruskal spacetime, there is a black hole region corresponding to $\{0 < r < 2M\}$, which is not there in the (flat) Rindler spacetime.

⁴We can avoid the different sign on $\{U = 0\}$ and $\{V = 0\}$ if we define the surface gravity on $\{V = 0\}$ with the opposite sign in (1.22). This is justified by the fact that on the portion of $\{V = 0\}$ and $\{U = 0\}$ to the future of the bifurcation surface, ξ^a necessarily has opposite directions.

1.3 Killing horizons

We have showed by analysis of the analytic metric that the event horizon of the Schwarzschild black hole is a bifurcate Killing horizon. However, the relation runs deeper: it is possible to show that the future event horizon of an asymptotically flat stationary black hole with $\kappa \neq 0$ is always a portion of a bifurcate Killing horizon, in presence of “physically interesting” matter.⁵ Indeed, this holds for all the known black hole solutions with $\kappa \neq 0$, and in particular for the Kerr–Newman black hole, which is the unique black hole solution in four-dimensional Einstein–Maxwell theory with a Killing vector that is timelike in a neighbourhood of asymptotic infinity (that is, it’s *stationary*). The other (very interesting!) possibility is a *degenerate Killing horizon*, defined as one with $\kappa = 0$.

This quite general characterization of black hole event horizons allows us to identify properties of these null hypersurfaces that don’t rely on knowing the analytic solutions, and in fact hold much more broadly than within general relativity. They are often referred to as “laws of black hole mechanics,” from a 1973 paper by Bardeen, Carter and Hawking [BCH73]. The precise statements of the laws vary somewhat depending on assumptions and setups, but the overarching message is that even at the level of classical physics, black holes exhibit a striking analogy with an ordinary thermal system. Such analogy is non-sensical at the level of classical physics, since by definition a black hole absorbs radiation but never emits it: *«the effective temperature of a black hole is absolute zero. [...] a black hole can be said to transcend the second law of thermodynamics»* [BCH73]. As we will see in the next section, things are very different once we include quantum effects.

Zeroth law *The surface gravity is constant on a bifurcate Killing horizon.*

Notice that this statement requires the change in sign mentioned in footnote 4. To prove this, one first uses Frobenius theorem to find an expression for κ^2

$$\kappa^2 = -\frac{1}{2}\nabla^a \xi^b \nabla_a \xi_b|_{\mathcal{N}}, \quad (1.25)$$

⁵The proper way to say this is to assume that the matter with stress-energy tensor T_{ab} satisfies the *dominant energy condition*, that is, $-T^a_b V^b$ is a future-directed causal or zero vector for any future-directed timelike vector V^a . This guarantees that an observer along V^a would not measure a spacelike energy-momentum current $-T^a_b V^b$.

from which we can show that the variation of κ along a vector field t^a that is tangent to \mathcal{N} is

$$\begin{aligned}\kappa t^c \nabla_c \kappa &= -\nabla^a \xi^b t^c \nabla_c \nabla_a \xi_b|_{\mathcal{N}} \\ &= -\nabla^a \xi^b t^c R_{bacd} \xi^d|_{\mathcal{N}}\end{aligned}\tag{1.26}$$

Here we used the so-called Killing vector lemma. Choosing $t^a = \xi^a$ shows that κ is constant along an orbit of ξ^a in \mathcal{N} , so it's constant along each generator of \mathcal{N} . Moreover, κ is also constant on the bifurcation surface: if we restrict to the bifurcation surface, and choose t^a to be a tangent vector, the derivative still vanishes because by definition ξ^a vanishes on the bifurcation surface. So, κ is constant everywhere on \mathcal{N} [KW91].

Note that the constancy of κ follows directly from the geometry of bifurcate Killing horizons, so it will apply to any horizon in any theory, provided it can be extended to a portion of a bifurcate Killing horizon. However, in the context of general relativity, it is also possible to prove directly that κ is constant on the (connected) future event horizon (not necessarily bifurcate) of a stationary black hole obeying the dominant energy condition [BCH73].

First law We can formulate a statement in any theory expressed by a Lagrangian \mathcal{L} that is only a functional of g_{ab} and other fields Φ , the Riemann tensor R_{abcd} , and symmetrized covariant derivatives of R_{abcd} and of the fields Φ . This is the form of the Lagrangian (if one exists) of any diffeomorphisms-invariant theory.

We begin with an asymptotically flat stationary black hole solution with a bifurcate Killing horizon, and consider a (not necessarily stationary), asymptotically flat solution of the linearized equations of motion around said solution. Then the following relation between quantities in the perturbation holds

$$\frac{\kappa}{2\pi} \delta S = \delta M - \Omega^i \delta J_i - \Phi^\alpha \delta Q_\alpha .\tag{1.27}$$

Here

- κ is the surface gravity of the bifurcate Killing horizon of the black hole
- S is defined as follows. Let Σ be the codimension-2 bifurcation surface, with binormal n_{ab} (that is, the volume element on the normal space to Σ in spacetime).

Then

$$S \equiv -2\pi \int_{\Sigma} \frac{\delta \mathcal{L}}{\delta R_{abcd}} n_{cd} \epsilon_{abc_3 \dots c_n} ,\tag{1.28}$$

where we have taken the functional derivative of \mathcal{L} with respect to R_{abcd} holding fixed all the other fields (including the metric).

- M , J_i and Q_α are, respectively, the ADM mass, the ADM angular momenta, and the electric charges, of the stationary black hole solution, and Ω^i and Φ^α are the conjugate angular velocities and electrostatic potentials of the horizon.

We will not prove this form of the first law, which is due to Iyer and Wald [IW94], but limit ourselves to a few comments.

First, note that we did not need to know anything about the analytic form of the solutions, apart from the fact that they solve the equations of motion, and their geometry (that is, we assume that we are discussing black holes with a bifurcate Killing horizon). Moreover, the statement of the first law is a highly non-trivial relation between quantities measured at the horizon (namely, κ and S) and quantities measured at asymptotic infinity (namely, the ADM and electric charges). These two observations go hand-in-hand. The way to prove the theorem is realizing that in all diffeomorphism-invariant theories there is an exact $(n - 1)$ -form that can be integrated over a hypersurface extending from the bifurcation surface to asymptotically flat infinity (its existence is guaranteed by the assumptions on the geometry). We then use Stokes' theorem to compute the vanishing integral, thus relating the contribution from the internal boundary (the bifurcation surface) and that from the asymptotic $(n - 2)$ -sphere, which is then expressed in terms of asymptotic conserved charges.

One highly non-trivial step in the proof is recognizing the contribution from the bifurcation surface in terms of variations of S in (1.28), the so-called Wald entropy. For general relativity, we have the Einstein–Hilbert action

$$\mathcal{L} = \frac{1}{16\pi} R \text{vol} = \frac{1}{16\pi} g^{ac} g^{bd} R_{abcd} \text{vol}, \quad (1.29)$$

from which

$$S = -\frac{1}{8} \int_{\Sigma} g^{ac} g_{bd} n_{cd} \epsilon_{abc_3 \dots c_n} = \frac{1}{4} \int_{\Sigma} \text{vol}_{\Sigma} = \frac{1}{4} A_{\text{h}}, \quad (1.30)$$

which, as we will review in section 2.1, is the famous expression for the entropy of black holes due to Bekenstein–Hawking.

Second law In contrast to the First law, the formulation of the Second law is much more restricted. In the context of four-dimensional general relativity, it also goes under the name of Hawking's area law [HE23].

In a strongly asymptotically predictable spacetime (i.e., provided we have sufficient control over the time evolution) satisfying the Einstein equations with matter satisfying the null energy condition (i.e. with “reasonable” matter), the area of the horizon does not decrease in time.

In order to generalize this result to more general theories that include derivatives or powers of the curvature, we would need to find a local functional of the geometry of the horizon that is non-decreasing in time evolution and for stationary spacetime reduces to the Wald entropy (1.28). As it turns out, this is a tall order: for instance, the definition (1.28) could suffer from ambiguities [JKM93], and even resolving them results in a non-decreasing functional only for linear perturbations [Wal15] or in a restricted sense [DR23].

Third law In the original paper [BCH73], there was also a conjectural third law of black hole mechanics, which, in the later formulation by Israel, stated

A subextremal black hole cannot become extremal in finite time by any continuous process, no matter how idealized, in which the spacetime and matter fields remain regular and obey the weak energy condition.

Quite recently, this conjecture was shown to be false by Kehle and Unger, who constructed spherically symmetric solutions to Einstein–Maxwell theory with a massless charged scalar that are Schwarzschild near the horizon for a period of advanced time, and then evolve to be exactly an extremal Reissner–Nordström black hole in a finite amount of time [KU22]. It is still possible to rule out these counterexamples, and thus maintain a third law in the formulation above, if one assumes that the matter stress-energy tensor is constrained by a stronger condition than the dominant energy condition, which takes the form of a bound on the charge to mass ratio [Rea24].

There is another statement of the third law of thermodynamics, which states that as the temperature goes to zero, the entropy goes to a universal constant determined by the degeneracy of the ground state of the system. Understanding how to make sense of this statement in the context of black holes requires taking into account quantum effects, and therefore it is beyond the scope of this section.

2 Hawking radiation and black hole thermodynamics

2.1 The laws of black hole thermodynamics

Classical results on the mathematical theory of black holes in $4d$ imply that gravitational collapse of an isolated body, though temporarily a messy and complicated process, where this body rotates, pulsates and throws matter in the universe, settles into a stationary black hole configuration described by the Kerr–Newman solution. Therefore, independently of the (potentially quite complicated) initial state, the final state of the system on and outside the event horizon is completely described by three physical quantities (M, J, Q) : mass, angular momentum and electric charge.

This is quite puzzling, as it seems to contrast with thermodynamics. Take an object with entropy (say a container filled with gas) and throw it in the black hole: after the object has crossed the event horizon, no signal from it can reach us far from the black hole, and thus the entropy of the universe has effectively been lowered.

This contradicts the Second Law of Thermodynamics, which is not good. This concern was shared by Bekenstein who came up with a revolutionary solution to the problem: black holes themselves have an entropy, and the total entropy of the black hole and of the universe outside the event horizon does not decrease [Bek72, Bek73].

What should the entropy of the black hole be? Bekenstein’s idea was to look at information theory. Loosely speaking, in information theory, the entropy of a system is a measure of lack of information about its internal configuration. Analogously, the entropy of a black hole is a measure of the inaccessibility of information to an outside observer about the internal configurations of the black hole. Thus, given a choice of mass, angular momentum and electric charge, the entropy of the black hole is a measure of the size of its equivalence class. In hindsight, it measures the number of quantum mechanical microstates, but this was not clear at all at the time.

Moreover, Bekenstein observed that Hawking had already proved that there is a quantity characteristic of a black hole that can never decrease: the area of its horizon (cf. the [second law of black hole mechanics](#)). Therefore, guided again by the analogy with the Second Law, he suggested that the entropy of a black hole should be proportional to its area. He even pointed out that using only general relativity (that is Newton’s constant) we cannot construct a combination with dimensions of length square that would allow us to obtain the (dimensionless) entropy, and we have to resort to quantum

mechanics, obtaining the remarkable prediction that a black hole with horizon of area A_{h} would be associated to an entropy⁶

$$S_{\text{bh}} = \eta A_{\text{h}} \frac{c^3}{G\hbar}, \quad (2.1)$$

where η is a dimensionless number, unspecified at this stage, and the combination $\ell_P = \sqrt{G\hbar/c^3}$ is the Planck length.

Contemporaries immediately realized the importance of Bekenstein's observation, but were very skeptical of the physical interpretation.⁷ In fact, Bardeen, Carter and Hawking pushed even further the analogy between black hole mechanics and thermodynamics, writing down the first version of the laws introduced in section 1.3 [BCH73]. Specifically, the first law of black hole mechanics (1.27) for a 4d Kerr–Newman black hole, with Wald's entropy evaluated for the Einstein–Hilbert action as in (1.30), is

$$\delta M = \frac{\kappa}{8\pi} \delta A_{\text{h}} + \Omega \delta J + \Phi_e \delta Q, \quad (2.2)$$

where A_{h} is the area of the horizon, Ω is the angular velocity of the horizon and Φ_e is the difference of electrostatic potential measured between asymptotic infinity and the horizon.

Compare with the First Law of thermodynamics: in any process involving a closed system with energy E , entropy S and charges Q_i , their variations are related by

$$dE = T dS + \sum_i \mu_i dQ_i, \quad (2.3)$$

where T is the temperature and μ_i are the chemical potentials associated to Q_i . In order to push further the analogy between black hole mechanics and thermodynamics suggested by the Bekenstein's formula (2.1), we identify mass and energy of the black hole, we recall that indeed (J, Q) are conserved charges, and (Ω, Φ_e) are the conjugate variables, but then we should associate a temperature to the black hole⁸

$$T_{\text{bh}} = \frac{\kappa}{8\pi\eta} \frac{\hbar}{c}. \quad (2.4)$$

⁶In this expression, we reinstate c , G and \hbar , keeping $k_B = 1$, so S is dimensionless, since $[c] = LT^{-1}$, $[G] = L^3 M^{-1} T^{-2}$, $[\hbar] = ML^2 T^{-1}$.

⁷A whirlwind account of the events of the time is at the beginning of [Pag04].

⁸Here again we reinstate c , G , \hbar , knowing that κ is an acceleration, e.g. for Schwarzschild $\kappa = c^4/(4MG)$, and since $k_B = 1$, $[T_{\text{bh}}] = ML^2 T^{-2}$.

We also notice that this identification would be consistent also with the other laws of black hole mechanics: surface gravity is constant on a bifurcate Killing horizon, just like temperature is constant on a system in thermal equilibrium; and the third law [in the formulation of the original paper](#) is parallel to one of the formulations of the third law of thermodynamics.

However, as already remarked by [\[BCH73\]](#), «*a black hole cannot be in equilibrium with black body radiation at any non-zero temperature, because no radiation could be emitted from the black hole whereas some radiation would always cross the horizon into the black hole.*» In fact, temperature is (always) a quantum effect (as shown in [\(2.4\)](#) by the presence of \hbar), so in this chapter we will look at the quantization of fields on a curved background without Poincaré isometries. As it turns out, this is quite subtle, and will eventually lead us to the proof that black holes *are* indeed thermodynamical objects emitting with a blackbody spectrum with temperature [\[Haw74\]](#)

$$T_H = \frac{\kappa}{2\pi} \frac{\hbar}{c}, \quad (2.5)$$

This also fixes the constant $\eta = 1/4$ in [\(2.1\)](#), consistently with [\(1.30\)](#). Knowing this, we can interpret the four laws of black hole mechanics reviewed in [section 1.3](#) as truly laws of black hole thermodynamics.

2.2 Quantum field theory on curved spaces

In order to show that black holes have a temperature, we need to investigate the behaviour of quantum field theory near the horizon. The issue is that this requires studying quantum field theory on a curved background, which is subtle.

When one first studies the quantization of a scalar field, one does so by considering a basis of plane wave solutions of the Klein–Gordon equation that have *positive frequency*, that is they are functions u_p such that

$$i\mathcal{L}_{\partial_t} u_p = \omega u_p, \quad \omega > 0. \quad (2.6)$$

On the space of positive-frequency solutions, one can introduce an inner product that is non-degenerate, Hermitian and positive-definite. One then expands a scalar field on orthonormal elements u_p and $\overline{u_p}$, and promotes the coefficients of the expansion to operators, a_p and a_p^\dagger , which satisfy the algebra of the creation and annihilation operators. The Hilbert space of the theory is then defined to be the symmetric Fock

space: the vacuum $|0\rangle$ is defined to be the state annihilated by all $a_{\mathbf{p}}$, and the other states are constructed by successive applications of $a_{\mathbf{p}}^\dagger$.

One immediate problem with the generalization of this approach to a curved background is that, even assuming a good causal structure (global hyperbolicity), there is no guarantee that there is a globally timelike Killing vector that could play the role of $(\partial_t)^a$ in (2.6) and define the positive-frequency subspace. But this is needed to define the annihilation operators, so there is no well-defined notion of “vacuum.” Therefore, there is no sensible interpretation of the states of the Hilbert space as “containing a fixed number of particles.” This looks like a radical departure from the case of flat space, but recall that we are doing quantum field theory. This is of course related to the concept of “particle,” but as we see not equivalent. In fact, even on flat space it would be wrong to interpret the “vacuum” as the state characterized by the absence of fluctuations, and “particle” may not be a useful concept. Talking about the presence of a particle also requires talking about the state of motion of the detector, because the mode decomposition (and consequently the notion of vacuum and particles) is global in nature. The canonical plane waves are agreed upon by all inertial detectors, who will thus also agree on a definition of the vacuum. However, this is not true for accelerating observers.

A better notion would be provided by a local quantity, such as the stress-energy tensor $\langle T_{ab}(x) \rangle$: if $\langle T_{ab}(x) \rangle = 0$, then the canonical transformations of tensors following diffeomorphisms would leave this invariant, and thus different observers would agree on the result.

In a sense, this shows that we’re using the wrong approach. Just as there is no preferred coordinate system in general relativity, where coordinate systems are irrelevant, so we conclude from our observations that there is no preferred Hilbert space of states when discussing quantum fields on spacetime. The corresponding mathematical statement is the Stone–von Neumann theorem: for systems with finite number of degrees of freedom there is a unique way (modulo unitary equivalence) of representing the canonical commutation relations on a Hilbert space, but this is not true for systems with an infinite number of degrees of freedom. In flat spacetime, this issue is avoided because Lorentz symmetry selects a preferred representation.

There is a way of introducing and discussing general relativity in a manifestly diffeomorphism-invariant way (which is the way we all learnt it): is there a way of

presenting quantum field theory without referring to the construction of a Hilbert space? Yes, it's the algebraic approach, in which the key role is played by the algebra of observables, and states are defined by the value of each observables (rather than the number of particles). In a more mathematical rephrasing, a quantum field theory is an assignment of an algebra of observables to each subset of spacetime, and a state is a linear map from the algebra of observables to \mathbb{C} that is normalized and positive. Notice that indeed the Hilbert space realization of the algebra of observables is not relevant to the discussion, though we can construct one (following an approach first written by Gelfand, Naimark and Segal).

We will not delve into algebraic quantum field theory, because it's too sophisticated for the questions we want to address. Instead, in the next section we will focus on the properties that define a thermal from the expectation value of the two-point function. However, we note that algebraic quantum field theory has been successful in providing rigorous proofs of the Unruh and Hawking effect for free theories, and establishes a mathematically sound framework to discuss questions such as the definition of the entropy of quantum field theories. It is once again an active field of research, after some years of dormancy. To get a feeling for this approach, in addition to the book [Wal95], one can read [HW14, Wit21].

2.2.1 The KMS condition

In the first courses in QFT, one usually describes correlators computed in the vacuum state, which is a pure state, that is, a unique ray in the projective Hilbert space. A system with finite temperature, instead, is a statistical ensemble of pure states: it's a mixed state described (only) by a density matrix.

As a concrete example, we look at a system with Hamiltonian H , and assume that the energy is the only quantum number labelling pure states and allowed to vary, so we are describing a system with temperature T in the canonical ensemble. The probability that the system at temperature T is in a pure state with energy E is given by

$$p_E = \frac{e^{-\beta E}}{Z}, \quad (2.7)$$

where $\beta = 1/T$, and Z is the canonical partition function of the system

$$Z(\beta) = \sum_E e^{-\beta E} = \text{Tr} e^{-\beta H}. \quad (2.8)$$

The expectation value of any observable \mathcal{O} at a temperature T is given by Gibbs' formula, which writes it as a weighted average over the value of \mathcal{O} on all the pure states:

$$\langle \mathcal{O} \rangle_\beta = \sum_E p_E \langle E | \mathcal{O} | E \rangle = \frac{\text{Tr} (\mathcal{O} e^{-\beta H})}{\text{Tr} (e^{-\beta H})}. \quad (2.9)$$

Notice that in order for this to make sense, both numerator and denominator must be defined separately, which imposes restrictions on the spectrum of the Hamiltonian.⁹

Now consider two observables \mathcal{A} and \mathcal{B} that evolve according to the Heisenberg picture

$$\mathcal{A}_t = e^{iHt} \mathcal{A} e^{-iHt}, \quad \mathcal{B}_t = e^{iHt} \mathcal{B} e^{-iHt}. \quad (2.10)$$

We define their correlators at time t as

$$\begin{aligned} G_+^\beta(t, \mathcal{A}, \mathcal{B}) &\equiv \langle \mathcal{A}_t \mathcal{B} \rangle_\beta = Z^{-1} \text{Tr} (\mathcal{A}_t \mathcal{B} e^{-\beta H}), \\ G_-^\beta(t, \mathcal{A}, \mathcal{B}) &\equiv \langle \mathcal{B} \mathcal{A}_t \rangle_\beta = Z^{-1} \text{Tr} (\mathcal{B} \mathcal{A}_t e^{-\beta H}). \end{aligned} \quad (2.11)$$

This definition, and the following use of the properties of the trace, requires the trace to be convergent, even if the operator is unbounded, and this is guaranteed, for instance, by choosing a compact spatial manifold. These correlators are defined by the physics for $t \in \mathbb{R}$, but we can analytically extend them by defining a complex $z = t + it_E$ where t and t_E are both real, and using the Heisenberg picture

$$\begin{aligned} G_+^\beta(z, \mathcal{A}, \mathcal{B}) &= Z^{-1} \text{Tr} [e^{i(z+i\beta)H} \mathcal{A} e^{-izH} \mathcal{B}], \\ G_-^\beta(z, \mathcal{A}, \mathcal{B}) &= Z^{-1} \text{Tr} [\mathcal{B} e^{izH} \mathcal{A} e^{-i(z-i\beta)H}]. \end{aligned} \quad (2.12)$$

Requiring that the exponents have negative real parts restricts the domain of holomorphicity of these functions, namely they are holomorphic in

$$\begin{aligned} G_+^\beta(z, \mathcal{A}, \mathcal{B}) &\quad -\beta < \text{Im } z < 0, \\ G_-^\beta(z, \mathcal{A}, \mathcal{B}) &\quad 0 < \text{Im } z < \beta, \end{aligned} \quad (2.13)$$

and $G_\pm^\beta(t, \mathcal{A}, \mathcal{B})$ is their limiting value when $\text{Im } z \rightarrow 0^\mp$. In fact, more is true: provided $-\beta \leq \text{Im } z \leq 0$, we can use the cyclicity of the trace to show the functional relation

$$G_+^\beta(z, \mathcal{A}, \mathcal{B}) = G_-^\beta(z + i\beta, \mathcal{A}, \mathcal{B}). \quad (2.14)$$

⁹For completeness, we write the result also using the density matrix. This is an operator defined by

$$\rho = \frac{1}{Z} e^{-\beta H},$$

and thus (2.9) can be also written as $\langle \mathcal{O} \rangle_\beta = \text{Tr } \rho \mathcal{O}$.

This relation goes under the name of *KMS condition* (after Kubo and Martin–Schwinger). It is sometimes also written in on the real axis as

$$\langle \mathcal{A}_t \mathcal{B} \rangle_\beta = \langle \mathcal{B} \mathcal{A}_{t+i\beta} \rangle . \quad (2.15)$$

We can use the KMS relation (2.14) to construct a periodic function throughout the plane apart from the lines $\text{Im } z = \ell\beta$ with $\ell \in \mathbb{Z}$, defined by

$$\begin{aligned} \mathcal{G}^\beta(z, \mathcal{A}, \mathcal{B}) &= G_-^\beta(z, \mathcal{A}, \mathcal{B}) & 0 < \text{Im } z < \beta , \\ \mathcal{G}^\beta(z, \mathcal{A}, \mathcal{B}) &= G_+^\beta(z, \mathcal{A}, \mathcal{B}) & -\beta < \text{Im } z < 0 , \end{aligned} \quad (2.16)$$

and

$$\mathcal{G}^\beta(z, \mathcal{A}, \mathcal{B}) = G_+^\beta(z - i\ell\beta, \mathcal{A}, \mathcal{B}) = G_-^\beta(z - i(\ell - 1)\beta, \mathcal{A}, \mathcal{B}) , \quad (2.17)$$

for an appropriate integer ℓ . However, we still don't know about its analyticity on the real axis. On the other hand, if we preserve the causality requirements of relativistic quantum field theories, then $\mathcal{A}_t \mathcal{B} = \mathcal{B} \mathcal{A}_t$ for t in some open interval on the real axis, then $G_\pm^\beta(t, \mathcal{A}, \mathcal{B})$ are equal on that interval, and we can use the edge-of-the-wedge theorem¹⁰ to prove that $G_\pm^\beta(z, \mathcal{A}, \mathcal{B})$ are analytic continuations of each other and thus define a single holomorphic function on a connected region of the complex plane (excluding parts of the lines $\text{Im } z = \ell\beta$ with $\ell \neq 0$). Thus, in thermal relativistic quantum field theories (even interacting ones), using the Gibbs' formalism we can construct a single holomorphic function that is periodic in imaginary time, as guaranteed by the KMS condition.

In fact, in the algebraic approach to quantum field theory this reasoning is turned on its head, and the KMS condition is taken to be the defining property of the state of thermal equilibrium at a temperature $1/\beta$.

2.2.2 Unruh effect

As in the previous chapter, in order to appreciate subtleties of curved space, we begin with free scalar quantum field theory on flat space, eventually taking the viewpoint of the accelerated observer introduced in section 1.1.

¹⁰This is a generalization of Morera's theorem applied to contours passing through the “window” $\{\text{Im } z = 0, |\text{Re } z| < \sqrt{\langle \mathbf{x}, \mathbf{y} \rangle_h}\}$, and then breaking them along the real axis.

We want to construct an analytic extension of the propagators (or Wightman functions)¹¹

$$\begin{aligned} G_+^\infty(t, \mathbf{x}, \mathbf{y}) &\equiv \langle 0 | \phi(t, \mathbf{x}) \phi(0, \mathbf{y}) | 0 \rangle , \\ G_-^\infty(t, \mathbf{x}, \mathbf{y}) &\equiv \langle 0 | \phi(0, \mathbf{y}) \phi(t, \mathbf{x}) | 0 \rangle . \end{aligned} \quad (2.18)$$

We assume that the theory satisfies the causal requirements of relativistic quantum field theory, so the commutators of the fields at spacelike separation vanishes, and we have

$$G_+^\infty(t, \mathbf{x}, \mathbf{y}) = G_-^\infty(t, \mathbf{x}, \mathbf{y}) \quad |t| < |\mathbf{x} - \mathbf{y}| , \quad (2.19)$$

therefore, we can repeat the discussion outlined in the previous section, and construct a holomorphic function $\mathcal{G}^\infty(z, \mathbf{x}, \mathbf{y})$ such that

$$\mathcal{G}^\infty(z, \mathbf{x}, \mathbf{y}) = \begin{cases} G_-^\infty(z, \mathbf{x}, \mathbf{y}) & \text{Im } z > 0 \\ G_-^\infty(z, \mathbf{x}, \mathbf{y}) = G_+^\infty(z, \mathbf{x}, \mathbf{y}) & \text{Im } z = 0, \quad |\text{Re } z| < \sqrt{\langle \mathbf{x}, \mathbf{y} \rangle_h} \\ G_+^\infty(z, \mathbf{x}, \mathbf{y}) & \text{Im } z < 0 \end{cases} \quad (2.20)$$

and there may be branch cuts along the real axis for $|\text{Re } z| > |\mathbf{x} - \mathbf{y}|$. This is in fact a function only of a continuation of the Lorentzian distance between the two points, that is $F(-z^2 + |\mathbf{x} - \mathbf{y}|^2)$, and $F(w)$ is a holomorphic function of w except where $w = 0$, that is, at $z^2 = |\mathbf{x} - \mathbf{y}|^2$.

Now consider the correlator as measured by the accelerated observer introduced in section 1.1: she follows an orbit of the boost generator b^a . If we are in dimension greater than 2, we align the plane (t, x^1) to the boost, and we introduce Rindler coordinates on the right wedge of the plane, as in (1.15), via

$$t = \rho \sinh \alpha \eta, \quad x^1 = \rho \cosh \alpha \eta. \quad (2.21)$$

Then, the boost generator has the form $b = \partial_\eta$, and we leave untouched the coordinates that may be transverse the boost plane, that is \mathbf{x}_\perp . In these coordinates the flat space Wightman function becomes

$$\mathcal{G}^\infty(t, \mathbf{x}, \mathbf{y})|_{\text{Rindler coords}} = F(\rho_x^2 + \rho_y^2 - 2\rho_x \rho_y \cosh \alpha \eta + |\mathbf{x}_\perp - \mathbf{y}_\perp|^2). \quad (2.22)$$

¹¹The superscript ∞ refers to the fact that they are computed at zero temperature, that is, at infinite $\beta = 1/T$. In the next section we shall introduce a label β .

We want to show that the resulting function is a holomorphic function with periodicity β that is the analytic continuation of the propagators of a scalar field on Rindler space, that is, we want to identify the right-hand side of the equation above with $\mathcal{G}_{\text{Rin}}^\beta(\eta, (\rho_x, \mathbf{x}_\perp), (\rho_y, \mathbf{y}_\perp))$ and measure β . To do so, we need to perform an analytic continuation to the complex plane, show that the resulting function has the appropriate analytic structure (and in particular periodicity along imaginary translations) and that its values along the imaginary axis are the Green's function for the Wick rotation of the Klein–Gordon propagator. This latter property would define it uniquely, because the operator $(\partial_{t_E}^2 + \sum_i \partial_{x_i}^2 - m^2)$ is elliptic, and its Green's function is unique. This is in contrast to the Lorentzian Klein–Gordon operator, which is hyperbolic, and there are multiple Green's functions corresponding to different boundary conditions (e.g. Feynman's, “retarded”, “advanced”).

We first extend (2.22) to the complex plane introducing $\zeta = \eta + i\eta_E$, and we observe that this extension has poles at the spatial distance of the two points

$$\cosh \alpha \zeta = \frac{\rho_x^2 + \rho_y^2 + |\mathbf{x}_\perp - \mathbf{y}_\perp|^2}{2\rho_x \rho_y}, \quad (2.23)$$

and is periodic with period $\frac{2\pi}{\alpha}i$. Finally, we should show that the would-be $\mathcal{G}_{\text{Rin},E}^\infty(\eta_E, (\rho_x, \mathbf{x}_\perp), (\rho_y, \mathbf{y}_\perp))$ is the Green's function for the Wick rotation of the Klein–Gordon propagator. But notice that after Wick rotation, the Rindler coordinates (2.21) are just polar coordinates in a Euclidean two-dimensional space

$$t_E = \rho \sin \alpha \eta_E, \quad x^1 = \rho \cos \alpha \eta_E, \quad (2.24)$$

so the Wick-rotated differential operator is just a rewriting in polar coordinates of the flat Klein–Gordon operator, for which $\mathcal{G}_E^\infty(t, \mathbf{x}, \mathbf{y})|_{\text{polar coords}}$ is the Green's function (reading (2.22) right to left).

Therefore, we confirm that for the Rindler observer, the Minkowski vacuum state behaves like a thermal state with temperature

$$T_0 = \frac{\alpha}{2\pi}. \quad (2.25)$$

This result is not restricted to the free scalar theory we just looked at, but can in fact be generalized using algebraic QFT to an arbitrary interacting quantum field theory on flat space (a result proved by Bisognano and Wichmann).

As a matter of fact, this is *not* the actual temperature measured by the accelerated observer. Recall that the worldline of a Rindler observer has tangent vector $b = \partial_\eta$, so the normalized four-velocity according to the metric (1.15) is

$$u^a = \frac{1}{\alpha\rho} \left(\frac{\partial}{\partial\eta} \right)^a = \frac{1}{\alpha\rho} b^a. \quad (2.26)$$

On the other hand, the frequency ω above is measured using directly b^a . Therefore, the frequency measured by the observer is $\omega_{\text{obs}} = \omega/(\alpha\rho)$, leading to a Planck spectrum

$$\frac{1}{e^{2\pi\omega_{\text{obs}}\rho} - 1}, \quad (2.27)$$

and thus to the *Unruh temperature*¹²

$$T_U = \frac{\rho^{-1}}{2\pi} = \frac{|A|}{2\pi}, \quad (2.28)$$

where A is the magnitude of the proper acceleration of the observer. Notice that an observer at infinity, for which $\xi \rightarrow +\infty$, will find $T_U \rightarrow 0$, which is sensible, because the spacetime we are working on it's still just Minkowski.

The peculiar phenomenon we just discovered is named *Unruh effect*: in flat space, the vacuum defined by an inertial observer will be perceived by an accelerated observer as a *thermal* state with temperature (2.28). This is a physical effect: an accelerated observer with a particle detector will indeed detect particles. **[As you will see in exercise (), the Minkowski vacuum is not a pure state for the Rindler observer.]** However, there is no paradox or contradiction: the mode decomposition (and consequently the notion of vacuum and “particles”) is *global* in nature, and it requires knowledge of the observer’s history. A better notion would be provided by a *local* quantity, such as the stress-energy tensor $\langle 0_M | T_{ab} | 0_M \rangle$: if the inertial observer measures $\langle 0_M | T_{ab} | 0_M \rangle = 0$, then the stress-energy tensor measured by the accelerated observer, or any other observer related by diffeomorphism, is related by the induced tensor transformation, so it still vanishes: $\langle 0_M | T'_{ab} | 0_M \rangle = 0$.

¹²The fact that temperature changes with the observer is phenomenon often called Tolman (or Tolman–Ehrenfest) law: the local temperature T_{obs} measured by an observer travelling along the orbit of a timelike Killing vector k^a is such that $\sqrt{-k^2} T_{\text{obs}} = \text{const.}$ Here we labelled the constant T_0 . Note the analogy with (1.8).

2.3 Black holes

In the Wick-rotated Euclidean signature spacetime there is a geometric way of detecting the periodicity of the Green's function. The metric is

$$ds_E^2 = \alpha^2 \rho^2 d\eta_E^2 + d\rho^2 + d\mathbf{x}_\perp^2. \quad (2.29)$$

Focusing on the relevant \mathbb{R}^2 spanned by ρ and η_E , we see that this is locally isometric to \mathbb{R}^2 in polar coordinates when one identifies $\rho > 0$ as the radial distance and η_E as the angle, so the curvature vanishes. However, whether it's actually \mathbb{R}^2 depends on the identification of η_E . In \mathbb{R}^2 , taken a circle of radius r it's a fact that

$$\frac{\text{circumference}}{\text{radius}} = \frac{2\pi r}{r} = 2\pi. \quad (2.30)$$

Suppose we identify $\eta_E \sim \eta_E + \beta$, then in the Euclideanization of Rindler space we would have, for a circle of constant ρ_*

$$\frac{\text{circumference}}{\text{radius}} = \frac{\beta \alpha \rho_*}{\rho_*} = \beta \alpha. \quad (2.31)$$

Therefore, the geometry has a locally flat behaviour, and so it is smooth, if and only if we take $\beta = 2\pi/\alpha$, which is the same result obtained from the analysis of the Wightman functions! If this identification for the period of η_E is not made, the line element (2.29) is said to have a *conical singularity*. This is because indeed a cone can be obtained by cutting an angle from a flat plane and gluing together the two sides: the geometry would be smooth everywhere but at the apex of the cone, and parallel transport of vectors around the apex would result in a change of their orientation equal to the angle removed from flat space.

With this observation, we have connected the regularity of the Wick-rotated geometry with the periodicity of the (unique) Euclidean Wightman function and then, using the KMS condition, with the temperature. The thermal system on the Lorentzian geometry is mapped to the field theory on a Euclidean background with a circle direction. Isn't this amazing?

We can apply the same Euclidean method to a much more general class of solutions. The Unruh effect is essentially related to the bifurcate Killing horizon for the generator of boosts b^a in Minkowski flat space. However, one of the points of section 1 was that bifurcate Killing horizons share many properties, among which the fact that observers

along orbits of the generators are accelerated in an analogous way, with acceleration related to the surface gravity: do they all measure a temperature? Here we focus on a special case.

2.4 Near the horizon of a black hole

Let's look at a static spherically symmetric spacetime with line element

$$ds^2 = -f(r) dt^2 + \frac{dr^2}{f(r)} + r^2 d\Omega_2^2, \quad (2.32)$$

where $d\Omega_2^2$ is the line element of the round metric on S^2 . The Schwarzschild solution (1.16) falls in this class, as do many others. There is an obvious Killing vector $k = \partial_t$, with norm $k^2 = -f(r)$. We assume further that there is a largest r_+ such that $f(r_+) = 0$, and that $f(r) > 0$ for $r > r_+$, which is the region we focus on. So, k^a is timelike on our spacetime, and its norm vanishes on the surface $\mathcal{N} = \{r = r_+\}$. To further investigate \mathcal{N} , we need to introduce coordinates such that the metric is well-defined on \mathcal{N} , which are a generalization of the ingoing Eddington–Finkelstein coordinates.

Let $r_*(r)$ be the function such that

$$r'_*(r) = \frac{1}{f(r)}, \quad (2.33)$$

and define $v = t + r_*(r)$. In these coordinates, the line element has the form

$$ds^2 = -f(r) dv^2 + 2 dv dr + r^2 d\Omega_2^2, \quad (2.34)$$

which is still Lorentzian and non-degenerate at $r = r_+$. Moreover $k = \partial_v$. We can now compute

$$k_a|_{\mathcal{N}} = (dr)_a, \quad (2.35)$$

so it's normal to \mathcal{N} , which is then a Killing horizon for k^a . Furthermore, we find that

$$\nabla_a k^2 = -f'(r_+) k_a, \quad (2.36)$$

and comparing with (1.22), we identify the surface gravity of the horizon as

$$\kappa = f'(r_+)/2 \quad (2.37)$$

(with a normalization of the Killing vector still to be determined). Not only is this a Killing horizon: provided appropriate boundary conditions on $f(r)$ are imposed, this

would also be the event horizon of a black hole. In the following, we assume that $f'(r_+) > 0$, so that \mathcal{N} is a bifurcate horizon.

As with Rindler space, we perform a Wick rotation $t = it_E$ and study the geometry of the resulting Euclidean space

$$ds_E^2 = f(r) dt_E^2 + \frac{dr^2}{f(r)} + r^2 d\Omega_2^2. \quad (2.38)$$

For all $r > r_+$, the space above is smooth, but that's not necessarily true at $\{r = r_+\}$. Let's look at a neighbourhood of this set, where the line element can be well-approximated by

$$ds^2 \sim f'(r_+)(r - r_+) dt_E^2 + \frac{dr^2}{f'(r_+)(r - r_+)} + r_+^2 d\Omega_2^2. \quad (2.39)$$

We then introduce a new radial coordinate with origin at the horizon

$$\rho^2 = \frac{4}{f'(r_+)}(r - r_+). \quad (2.40)$$

Using this coordinate, the line element in a neighbourhood of the horizon looks like

$$ds^2 = \frac{f'(r_+)^2}{4} \rho^2 dt_E^2 + d\rho^2 + r_+^2 d\Omega_2^2. \quad (2.41)$$

But this is just the Euclidean Rindler metric (2.29)! So, we already know how to deal with the smoothness requirement: we should view ρ as a radial coordinate $\rho > 0$ and, to avoid conical singularities, we should identify

$$t_E \sim t_E + \beta, \quad \beta = \frac{4\pi}{f'(r_+)} = \frac{2\pi}{\kappa}, \quad (2.42)$$

which guarantees that the circles of constant ρ shrink smoothly as $\rho \rightarrow 0$. The resulting geometry is a product of a disc \mathbb{R}^2 and a 2-sphere, with a smooth metric. The disc is represented in figure 4.

In the meantime, we can appeal again to the chain of reasoning introduced earlier: regularity of the Wick-rotated geometry near the Killing horizon requires a periodicity for the adapted coordinate along the Killing vector, we identify this periodicity with the periodicity of the Euclidean Wightman function using the KMS condition, and then conclude that: an observer following an orbit of $k = \partial_t$ in the (Lorentzian) spacetime (2.32) is accelerated with acceleration

$$|A| = \frac{\kappa}{\sqrt{-k^2}} \sim \sqrt{\frac{f'(r_+)}{4(r - r_+)}} = \frac{1}{\rho},$$

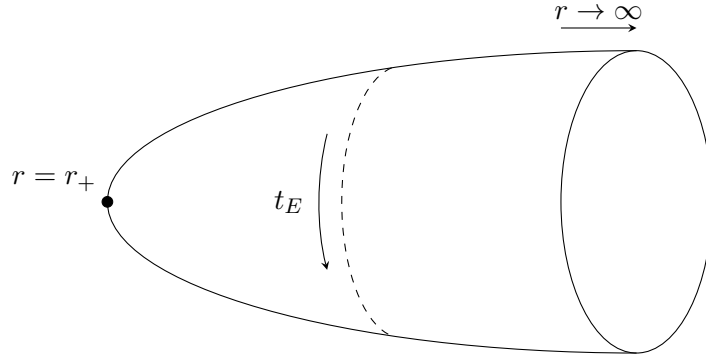


Figure 4: The topology of a static spherically symmetric black hole is $\mathbb{R}^2 \times S^2$. The metric on the \mathbb{R}^2 factor parametrized by (t_E, r) , and the ranges of the coordinates are such that the space caps off smoothly at $r = r_+$ with flat metric in a neighbourhood. The asymptotic behaviour as $r \rightarrow \infty$ depends on the cosmological constant: here we represent an asymptotically flat solution, where the \mathbb{R}^2 factor, as $r \rightarrow \infty$, becomes a cylinder with metric $dt_E^2 + dr^2$, since the size of the circles at constant r doesn't grow. This is the “cigar” geometry.

and she will detect a surrounding thermal bath of particles with temperature (2.28)

$$T = \frac{|A|}{2\pi} = \frac{\kappa}{2\pi\sqrt{-k^2}}, \quad (2.43)$$

where $\sqrt{-k^2}$ in the denominator is the redshift factor due to the change to the observer's frame (see (1.8) and footnote 12). Suppose that k^a is now such that $k^2 \rightarrow -1$: then we find an interpretation for the constant in Tolman's law: it's the temperature measured by an observer far from the horizon, and in particular it's

$$T_H = \frac{\kappa}{2\pi}. \quad (2.44)$$

This is referred to as the *Hawking temperature*. Notice the difference with the previous section: if the bifurcate Killing horizon is actually the event horizon of a spacetime, the Hawking temperature is non-vanishing, whereas in Rindler space there was no preferred observer along b^a , and the “Hawking temperature” vanished (as commented right after (2.28)).

For future use, also note that the application of the Green's function (or regularity) method does not pick up the temperature measured by the observer (2.28), but rather

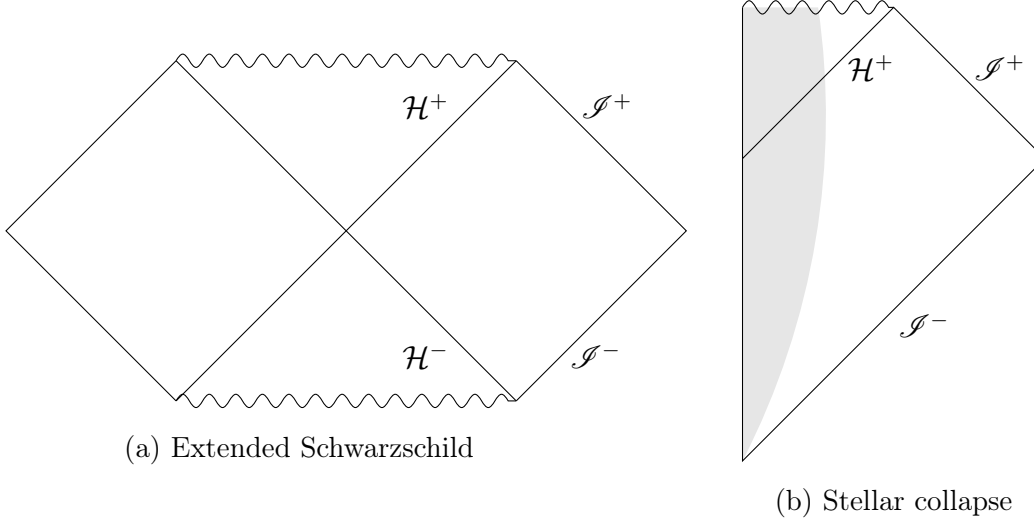


Figure 5: Penrose diagrams of (a) the extended Schwarzschild black hole, and (b) a spherically symmetric stellar collapse (in grey the star).

the constant T_0 (2.25) in Tolman’s law (see footnote 12), which we identified as the Hawking’s temperature.

We also recall that the Lorentzian Killing horizon \mathcal{N} is a codimension-1 set, and the Bekenstein–Hawking entropy (1.30) is given by the area of the bifurcation surface. After the Wick rotation and the construction of the regular Euclidean manifold, the bifurcation surface, where the Killing vector vanishes, has become the 2-sphere over the origin of the disc, that is, the locus $\{r = r_+\} \times S^2$. In fact, that’s what remains of the entire Killing horizon.

2.4.1 Far from the horizon of a black hole

At this point one may ask whether we have already reached our goal of proving the *Hawking effect*: a black hole with surface gravity κ radiates particles with a thermal spectrum at temperature (2.44). We have not.

First, we have not because in order to have a bifurcate horizon with the right analyticity properties in, say, the Schwarzschild black hole, we would need the entire conformal diagram 5a, but in realistic black holes originated from stellar collapse part of the diagram is hidden due to the presence of matter 5b. Therefore, we need to study a different situation. Second, in applying the analysis of the Unruh effect to curved spacetime with a Killing horizon we have made an assumption about the existence of

a vacuum state. In fact, a more careful analysis (see [Wal95, sec. 5.3]) shows that in order to generalize the Unruh effect to curved spacetime one requires the existence of a vacuum state invariant under the isometry generating the horizon and non-singular, in the sense that the expectation value of the stress-energy tensor in the vacuum should be non-singular. Such a state does not exist if the Killing vector generating the horizon is not globally timelike: whilst this is not a problem for Schwarzschild, it is for the Kerr black hole. Therefore, there is no Unruh effect for the Kerr black hole. However, an observer far from the Kerr black hole would still see particle production.

We will not review the original computation by Hawking [Haw75], which is highly non-trivial and which the interested reader can find reviewed in detail in [Rea20]. One crucial aspect of the computation involving spherically symmetric gravitational collapse represented in figure 5b is that, even though the metric outside the matter is the *static* Schwarzschild metric (because of Birkhoff’s theorem), the geometry is not overall stationary, because the metric inside the matter is not. Therefore, observers on \mathcal{I}^+ and \mathcal{I}^- would not agree on the definition of the “vacuum.” In particular, consider the propagation of radiation from \mathcal{I}^- in the collapsing matter and then scattering to \mathcal{I}^+ leads to the conclusion that in the state that the observer on \mathcal{I}^- calls “vacuum,” the observer on \mathcal{I}^+ will measure a spectrum of particles that at late retarded times only depends on the surface gravity of the black hole.

Differently from the Unruh effect, the final result for the number of particles measured by the late-time observer in the early-time vacuum will not be a purely Planckian spectrum at the Hawking temperature (2.44), but instead will also include a “grey-body” factor, which takes into account the fact that the black hole emits *and* absorbs radiation. However, the ratio of absorption and emission rates is independent of the greybody factor, signalling that the black hole with surface gravity κ is in equilibrium with a heat bath at temperature $T_H = \kappa/2\pi$, as in (2.44).

The computation can be generalized to non-scalar field theories, and to different gravitational backgrounds. For instance, we could have also studied a collapsing back with angular momentum and charge. After a dynamical state, the black hole would settle down to a stationary Kerr–Newman solution, which is described only by its mass, angular momentum and charge. At late times, the black hole is in equilibrium with a heat bath with temperature given again by T_H in (2.44) (clearly with the appropriate

surface gravity), and the emitted particles preferably have angular momentum and charge with the same sign as the black hole itself (this is the same that one would expect for a rotating charged black body) [Haw75].

2.5 Where to now?

We argued that in presence of a gravitational collapse leading to a stationary black hole, at late time an observer would measure an outgoing flux of particles distributed along the spectrum according to Planck’s law with a greybody factor and a temperature T_H (2.44). In order to get to this result, we needed to consider the behaviour of quantum fields near the black hole, but in the semiclassical approximation, that is, neglecting the quantization of the gravitational field itself, and the backreaction of the fields on the geometry. Before moving on with the topics, we stop to make some comments.

It is useful to have an intuitive picture of the origin of the black hole radiation. The Hawking effect has an analogue in (flat space) electrodynamics, which had been known for some time, the Schwinger effect [Sch51]. Recall that the quantum field theory vacuum is not really empty, and pairs of particle-anti particle are continuously created. For instance, let’s focus on electron-positron pairs and apply a strong electric field to a region of (supposedly) empty space: as the pair of particles is created, the field pulls the electron in one direction and the positron in the opposite direction. If the field is sufficiently strong, we will be able to observe the creation of *real* electrons and positrons at the opposite ends of the region, and there will be a current flowing.

The picture with gravity is slightly different, as all particles now have the same “charge.” In this case, the crucial role is played by the black hole region of spacetime created by the strong gravitational field: one member of the electron-positron pair could fall inside the horizon, and the other could flow off to infinity, having now become a *real* particle measured by the far observer.

As already anticipated at the beginning of section 2.1, we have showed that that black holes are indeed thermodynamical objects. Classical thermodynamics can be derived, using statistical physics, from quantum microstates. For a black hole, in order to reproduce the Bekenstein–Hawking entropy, we would need to describe $N \sim \exp(A/4)$ microstates. This requires a quantum theory of gravity. One of the most impressive successes of string theory as a theory of quantum gravity has been the

counting of the degeneracy of states corresponding to a microscopic description of a supersymmetric black hole, reproducing the Bekenstein–Hawking entropy, and even providing the leading correction to the result [SV96].

The idea that black holes must be assigned an entropy was based on the Second Law of Thermodynamics, and the suggestion was that in presence of a black hole, this would be modified to a *Generalized Second Law*: in any physical process

$$\Delta(S_{\text{matter}} + S_{\text{bh}}) \geq 0. \quad (2.45)$$

Here, recall that $S_{\text{bh}} = A/4$ is entropy of the black hole, which is just a function of (M, J, Q) , so it's obvious to the entropy of the matter/radiation falling in the black hole and it is not at all obvious that it would increase sufficiently to compensate the decrease in S_{matter} . In particular, requiring that the Generalized Second Law must hold would also imply a bound on S_{matter} as a function of its extensive parameters. Such a bound is intrinsically related to gravity, as it has at its core the tenet that one cannot fill a region of space with an arbitrarily high number of degrees of states without encountering a gravitational instability and forming a black hole. In particular, this leads to the idea of a holographic principle: *the physics in a region with boundary area A is fully described by no more than $A/4$ degrees of freedom*. This suggestion is in stark contrast with the predictions of local field theory, and it should be a property of a theory of quantum gravity. The most concrete instance of the holographic principle is the realization of the AdS/CFT correspondence in string theory (see [Bou02] for a review).

In the derivation above, we ignored the backreaction of the radiation on the geometry. A proper study of the backreaction requires a quantum theory of gravity, but from the presence of radiation itself, one can immediately draw some puzzling conclusions about unitary evolution. Consider matter starting its collapse in a pure (definite) quantum state. It will form a black hole, which will then radiate and eventually evaporate completely, leaving only its thermal radiation in the universe. The latter state can now only be described using a density matrix, it's a mixed state. Thus, we seem to have described the evolution from a pure to a mixed quantum state, which would not be consistent with unitary evolution in quantum mechanics. This is roughly the content of the *information paradox*: does the black hole evolve unitarily in time?

It is fair (and fun) to say that the subject of the evolution of a thermal black hole is still a heated subject, and we will not cover it. For some reviews see, for instance, [Mat09, AHM⁺20]. We limit ourselves to notice that the previous comments on string theory realizations imply that black holes do indeed evolve unitarily.

Finally, to conclude this section, let's view some numbers. For a static black hole with mass M

$$\begin{aligned} T_H &= \frac{\kappa}{2\pi} \frac{\hbar}{k_B c} = \frac{1}{8\pi M} \frac{\hbar c^3}{G k_B} \sim 6.17 \cdot 10^{-8} \cdot \frac{M_\odot}{M} \text{ K}, \\ S_{\text{bh}} &= \frac{A}{4} \frac{k_B c^3}{G \hbar} = 4\pi M^2 \frac{G k_B}{c \hbar} \sim 1.05 \cdot 10^{77} \cdot \frac{M^2}{M_\odot^2} k_B. \end{aligned} \tag{2.46}$$

Therefore, modelling with crude approximation an astrophysical black hole ($M \sim 10^6 M_\odot$) with a Schwarzschild solution, we find that its Hawking temperature is minuscule, much lower than the temperature of the CMB radiation, whereas its entropy is enormous, much higher than the entropy of matter in the same volume.

3 The gravitational path integral

In this section we look at a naïve but insightful approach to the study of quantum aspects of gravity: the gravitational path integral.

3.1 Definition

Let’s begin by recalling the definition of the path integral in quantum mechanics. This is applied to the computation of the amplitude that a particle at position x_1 at time t_1 is found at position x_2 at a later time t_2 . Feynman’s idea is to compute this by the evaluation of the integral over all trajectories between x_1 and x_2 , weighted by an oscillatory contribution due to the classical action:¹³

$$\langle x_2 | e^{-it\frac{H}{\hbar}} | x_1 \rangle = \int_{x=x_1}^{x=x_2} \mathcal{D}x e^{i\frac{S}{\hbar}}, \quad (3.1)$$

where $S = \int_0^t \left(\frac{m}{2} \dot{x}^2 - V(x) \right) ds$. A far-reaching observation is that the time propagator $e^{-it\frac{H}{\hbar}}$ is naively related to the density matrix operator $e^{-\beta H}$ by $t = -i\beta\hbar$ (where $\beta = 1/k_B T$). As we saw and justified in some detail in the previous section, the relation is deeper and more involved, but one can in fact construct the path integral representation of the density matrix operator (see footnote 9), and view it as a “path integral representing evolution in imaginary time.” In fact, the construction of the “path integral representation” of $e^{-\beta H}$ can be done in a much more rigorous way than that of $e^{-itH/\hbar}$, even to the level of satisfaction of a mathematician (and indeed it is named Feynman–Kac formula after the mathematician Mark Kac). The result is formally represented by the expression

$$\langle x_2 | e^{-\beta H} | x_1 \rangle = \int_{x=x_1}^{x=x_2} \mathcal{D}x e^{-\frac{S_E}{\hbar}}, \quad (3.2)$$

where now S_E is the *Euclidean* action

$$S_E = \int_0^{t_E} \left(\frac{m}{2} \dot{x}^2 + V(x) \right) ds_E, \quad (3.3)$$

and $t_E = \beta\hbar$. Notice that in this variables, the relation between the time propagator and the density matrix has become the canonical Wick rotation $t = -it_E$, and indeed

¹³In this section we reintroduce k_B and \hbar .

this can be used to relate S and S_E

$$\begin{aligned}
iS &= i \int_0^t \left(\frac{m}{2} \left(\frac{dx}{ds} \right)^2 - V(x) \right) ds = i \int_0^{-it_E} \left(\frac{m}{2} \left(\frac{dx}{ds} \right)^2 - V(x) \right) ds \\
&= i \int_0^{t_E} \left(-\frac{m}{2} \left(\frac{dx}{ds_E} \right)^2 - V(x) \right) (-i ds_E) \\
&= -S_E
\end{aligned} \tag{3.4}$$

where in the second line we introduced $s = -is_E$. In higher dimensions, this procedure justifies the adjective “Euclidean”, but for now notice that S_E has the same form as the energy $T + V$. Finally, we can construct the canonical partition function $Z(\beta)$, which is obtained from (3.2) by identifying x_1 and x_2 and summing over them, thus obtaining a trace

$$Z(\beta) = \int \mathcal{D}x e^{-\frac{S_E}{\hbar}}, \tag{3.5}$$

where the sum is over all path that close after “imaginary time” β . We now see that the path integral picture of the trace naturally enforces the fact that the thermal partition function should be computed over Euclidean backgrounds where “time” is identified in a circle of radius β , which we discussed from the periodicity of the Green’s function in section 2.2.1. We have come full circle (pun intended)!

Guided by this, and the success of the path integral in quantum field theory, we want to extend it to gravity. We want to compute the probability amplitude of evolving from a state described by a metric g_1 on a spacelike surface Σ_1 with matter fields Φ_1 to a state described by metric g_2 on a spacelike surface Σ_2 with matter fields Φ_2 , which can be formally written as

$$\langle g_2, \Phi_2, \Sigma_2 | g_1, \Phi_1, \Sigma_1 \rangle = \int \mathcal{D}g \mathcal{D}\Phi e^{i\frac{S}{\hbar}}. \tag{3.6}$$

Here we introduced the (undefined) path integral measures $\mathcal{D}g$ and $\mathcal{D}\Phi$, and the integral is taken over all the field configurations satisfying the appropriate boundary conditions. More precisely, to avoid further divergences we have to assume that either $\Sigma_{1,2}$ are compact or that they are joined by a timelike tube, to make the overall region of spacetime we are concerned with compact. Thus we are working with a compact region of spacetime M with boundary $\partial M = \Sigma_1 \cup \Sigma_2 \cup C$, as in figure 6.

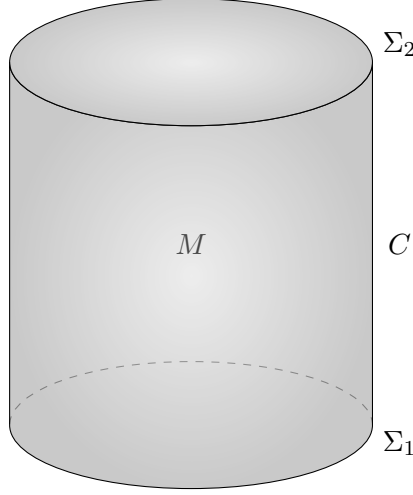


Figure 6: A schematic representation of the application of the gravitational path integral. To compute the amplitude (3.6) one should integrate over all configurations filling ∂M with the appropriate boundary conditions on ∂M .

3.2 Gravity action

The action of the metric in general relativity is the Einstein–Hilbert action

$$S_{\text{EH}} = \frac{1}{16\pi} \int (R - 2\Lambda) \sqrt{-g} \, d^4x, \quad (3.7)$$

to which we should add the Lagrangian describing the matter fields S_m coupled to the metric. If we tried to use $S = S_{\text{EH}} + S_m$, we would immediately find an issue: the Einstein–Hilbert action does not generically yield a well-posed variational problem on a space with a boundary! That is, varying the Einstein–Hilbert action does not only give the equations of motion, but also a boundary term. If the space is asymptotically flat and the metric and its derivatives decay sufficiently fast at infinity, this boundary term can be ignored, as it is usually done. In general, we cannot do that.

More precisely, varying S_{EH} leads to

$$\begin{aligned} \delta S_{\text{EH}} &= \frac{1}{16\pi} \int_M \left[\left(R_{ab} - \frac{1}{2} R g_{ab} + \Lambda g_{ab} \right) \delta g^{ab} + \nabla_a X^a \right] \sqrt{|g|} \, d^4x \\ &= \frac{1}{16\pi} \int_M \left(R_{ab} - \frac{1}{2} R g_{ab} + \Lambda g_{ab} \right) \delta g^{ab} \sqrt{|g|} \, d^4x + \frac{1}{16\pi} \int_{\partial M} n_a X^a \sqrt{|h|} \, d^3x, \end{aligned} \quad (3.8)$$

where we have used the divergence theorem in the second step, denoting by h_{ab} the induced metric on ∂M . The bulk term in (3.8) consists of the Einstein equations with

a cosmological constant, whereas the boundary term is given by

$$\begin{aligned} X^a &= g^{bc} \delta \Gamma_{bc}^a - g^{ab} \delta \Gamma_{bc}^c, \\ \delta \Gamma_{bc}^a &= \frac{1}{2} g^{ad} (\delta g_{db;c} + \delta g_{cd;b} - \delta g_{bc;d}), \end{aligned} \tag{3.9}$$

and at the end of the day its presence is due to the fact that the Riemann tensor is second order in the derivatives of the metric. In order to define a well-posed variational problem, we need the boundary term to vanish on its own (which requires both the metric and its derivatives to vanish on the boundary) or to add a further boundary term that cancels the variation of the Einstein–Hilbert action. Such a term would have to be a special combination of the first derivatives of the induced metric. Luckily, such a term does exist: it’s the *extrinsic curvature* of an embedded surface.

3.2.1 Gibbons–Hawking–York term

Consider a timelike or spacelike surface Σ with unit normal n_a (that is $n_a n^a = \pm 1$ depending on whether it’s timelike or spacelike), and induced metric $h_{ab} = g_{ab} \mp n_a n_b$. The latter is sometimes denoted first fundamental form and has the property that h^a_b is a projector on the tangent space to Σ (indeed $h^a_b n^b = 0$). The normal is defined on Σ , but we can extend it to the ambient space M as we want (the final result does not depend on the extension). We can then use the Lie derivative to measure how does the induced metric on Σ varies as we move along an integral curve of n^\sharp : the result is the extrinsic curvature

$$K_{ab} = \frac{1}{2} \mathcal{L}_{n^\sharp} h_{ab}, \tag{3.10}$$

which is a rank-2 symmetric tensor, sometimes denoted second fundamental form. We can also write it in a (non-trivially) equivalent way as

$$K_{ab} = h^c_a h^d_b \nabla_c n_d. \tag{3.11}$$

By its definition it is clear that K_{ab} depends also on the way the surface sits in the ambient space. In contrast, the *intrinsic curvature* of Σ is measured by the Riemann tensor $R[h]^a_{bcd}$ associated to h_{ab} (and to its Levi-Civita connection): it is related to the extrinsic curvature and the curvature of the ambient space $R[g]^a_{bcd}$ by the *Gauss equation*

$$R[h]^a_{bcd} = h^a_e h^f_b h^g_c h^h_d R[g]^e_{fgh} \pm 2K_{[c}^a K_{d]b}. \tag{3.12}$$

Finally, we also define the trace of the extrinsic curvature by $K \equiv h^{ab}K_{ab}$.

To get a feeling for the meaning of the extrinsic curvature, consider a surface Σ in M defined by the level set $f = 0$ with $f : M \rightarrow \mathbb{R}$ and $df \neq 0$ on Σ . The unit normal is $n_a = \mathcal{N}(df)_a$ where \mathcal{N} is a function on M such that n_a is a unit normal, and

$$\nabla_c n_b = \mathcal{N} \partial_{bc}^2 f + \partial_c \mathcal{N} (df)_b. \quad (3.13)$$

Using (3.11), the extrinsic curvature is

$$K_{ab} = h^c{}_a h^d{}_b \mathcal{N} \partial_{cd}^2 f. \quad (3.14)$$

Therefore, if $M = \mathbb{R}^d$, it is the projection to the tangent space to the surface of the second derivative of f . Since $f(p) = 0$ for $p \in \Sigma$, and the projection of the first derivative is vanishing (as the latter is proportional to the normal), K_{ab} is the leading approximation to f in a neighbourhood of p , once projected to the tangent plane, or the best approximation of the hypersurface by a paraboloid. It measures how much the hypersurface moves “away” from the tangent plane at a point in the direction of the normal to the point, and thus corresponds to our intuitive definition of “curvature.”¹⁴

As an example of computation of extrinsic curvature that will be particularly useful in a little while, consider a timelike hypersurface of constant r in the static spherically symmetric spacetime (2.32). The unit normal to such surface is $n = f(r)^{-1/2} dr$, the dual vector field is $n^\sharp = f(r)^{1/2} \partial_r$ and the induced metric is $h = -f(r) dt^2 + r^2 d\Omega_2^2$. We can then use the definition of the Lie derivative to compute K_{ab} using (3.10), which in a coordinate basis is

$$\begin{aligned} K_{\mu\nu} &= \frac{1}{2} (n^\rho \partial_\rho h_{\mu\nu} + h_{\mu\rho} \partial_\nu n^\rho + h_{\rho\nu} \partial_\mu n^\rho) = \frac{1}{2} \sqrt{f(r)} \partial_r h_{\mu\nu} \\ &= \sqrt{f(r)} \left(-\frac{f'(r)}{2} dt^2 + r g(S^2) \right)_{\mu\nu}, \end{aligned} \quad (3.15)$$

with trace

$$K = h^{\mu\nu} K_{\mu\nu} = \frac{f'(r)}{2\sqrt{f(r)}} + \sqrt{f(r)} \frac{2}{r}. \quad (3.16)$$

As stated earlier, adding a multiple of the extrinsic curvature of the boundary to S_{EH} allows us to construct a well-defined variational problem. More precisely, we add

¹⁴For instance, you can convince yourself that a cylinder has vanishing intrinsic curvature (you can create it by rolling up a flat piece of paper) but it has non-vanishing extrinsic curvature.

the *Gibbons–Hawking–York term*

$$S_{\text{GHY}} = \frac{1}{8\pi G} \int_{\partial M} K \sqrt{|h|} \, d^3x, \quad (3.17)$$

Together with this term, as you will prove in the problem sheet, the variation (3.8) changes to

$$\begin{aligned} \delta(S_{\text{EH}} + S_{\text{GHY}}) &= \frac{1}{16\pi} \int_M \left(R_{ab} - \frac{1}{2} R g_{ab} + \Lambda g_{ab} \right) \delta g^{ab} \sqrt{|g|} \, d^4x \\ &\quad + \frac{1}{16\pi} \int_{\partial M} \Pi_{ab} \delta h^{ab} \sqrt{|h|} \, d^3x, \end{aligned} \quad (3.18)$$

where

$$\Pi_{ab} = K_{ab} - K h_{ab}. \quad (3.19)$$

Therefore, the boundary term vanishes if we impose Dirichlet boundary conditions on the metric, that is, $\delta h^{ab} = 0$, or if we impose the Neumann boundary condition by fixing the derivative $\Pi_{ab} = 0$, and we are only left with the equations of motion, that is, with a well-posed variational problem. The combination (3.19) is referred to as *Brown–York stress-energy tensor*, for reasons that we’ll see later on in section 4.3.

3.2.2 Euclidean space

We have established that the action S appearing in (3.6) as the weight of each contribution to the gravitational path integral contains $S = S_{\text{EH}} + S_{\text{GHY}} + S_m$ (and potentially also fixed terms that could be extracted from the definition of the measure, as we shall soon see). An immediate problem, though, is that for Lorentzian metrics and unitary matter fields, S is real, so the integrand oscillates and the path integral will not generically converge. Relatedly, the boundary problem to be solved in order to find the configurations contributing to the path integral involves a hyperbolic equation, which, as you know from trying to find propagators in Lorentzian QFT, is not a well-posed problem, as the existence of the solution is not guaranteed and even so it is not unique (e.g. the non-uniqueness of the propagator for a real scalar field).

One solution to improve this state of affairs in stationary spacetime, as already discussed in the previous section, is to perform a Wick rotation, in which case we can define a unique vacuum by imposing the asymptotic boundary conditions for the fields. Moreover, the propagator found in this way agrees upon analytic continuation with the

Feynman's propagator. As already mentioned, this procedure for quantum mechanics leads to the Feynman–Kac formula, which mathematicians too are happy with.

It is reasonable to suggest that the same procedure be applied to the gravitational path integral (3.6). Letting t be the length in time of the timelike tube connecting the surfaces Σ_1 and Σ_2 , if we define $t = -it_E$, then the induced metric on the tube is now positive-definite and the gravitational path integral involves now the boundary problem of an elliptic differential equation. Namely, the integration is over all the positive-definite metrics g_E on the compact region M that induce the positive-definite metric h_E on ∂M . The weight of each configuration is now e^{-S_E} where

$$\begin{aligned} S_E &= -iS|_{\text{Wick rotated}} \\ &= -\frac{1}{16\pi} \int_M (R - 2\Lambda) \sqrt{g_E} d^4x - \frac{1}{8\pi} \int_{\partial M} K \sqrt{h_E} d^3x - S_{m,E}, \end{aligned} \quad (3.20)$$

and now the Ricci scalar R and the extrinsic curvature K are computed using g_E and h_E , respectively. These prescriptions are the essence of *Euclidean quantum gravity*. This approach to quantum gravity has been very successful, even (and surprisingly) beyond the explanation of the thermodynamic properties of black holes, though there are still some issues to be clarified, which we will review in section 4.3.

For the time being, we ignore all issues of convergence or subtleties in the definition, we put on blinders and compute. What should we begin with? As we already discussed at length, Euclidean field theory shines when it is applied to thermal Lorentzian systems: the description of a Lorentzian system on a background $\mathbb{R}_t \times \Sigma$ at a temperature T is mapped to a Euclidean problem on $S^1 \times \Sigma$, where the angular coordinate on S^1 has period $\beta = 1/k_B T$. The characteristic state function of the system is the free energy F , directly related to the partition function of the system via

$$\beta F(\beta) = -\log Z(\beta). \quad (3.21)$$

As reviewed in section 3.1, in the path integral approach the canonical partition function is given by the integral

$$Z(\beta) = \int \mathcal{D}g_E \mathcal{D}\Phi_E e^{-\frac{S_E}{\hbar}}, \quad (3.22)$$

where the integration is over all metrics and field configurations with the appropriate boundary condition on $\partial M \cong S^1 \times \partial\Sigma$. Note that this is a tantalizing reminder of the holographic nature of gravity, and will become better defined in AdS.

How do we approach the Euclidean gravitational path integral? The short answer is that we do not, because we do not know how to make sense of the measure over the space of metrics. Instead, we look at the *semiclassical approximation*: in the limit when \hbar is much smaller than the action of the typical path,¹⁵ the generalized contour integral would be (hopefully) dominated by the configurations $(\overline{g_E}, \overline{\Phi_E})$ corresponding to stationary points of the action S_E , which are the solutions to the classical equations of motion. In this limit, we obtain

$$S_E(g_E, \Phi_E) = S_E(\overline{g_E}, \overline{\Phi_E}) + S_E^{(2)}(g_E, \Phi_E) + \dots \quad (3.23)$$

where $S_E(\overline{g_E}, \overline{\Phi_E})$ is the on-shell action evaluated on the solutions of the classical equations of motion and it is commonly denoted I , $S_E^{(2)}(g_E, \Phi_E)$ is a functional quadratic in the perturbation around the classical solution, and we ignored higher order terms in the expansion of the action. So, the free energy in (3.21) is given by

$$\beta F = \frac{1}{\hbar} I - \log \int \mathcal{D}g_E \mathcal{D}\Phi_E e^{-\frac{S_E^{(2)}}{\hbar}} + \dots \quad (3.24)$$

The second term in the expansion is referred to as the *one-loop contribution*, and it represents the effects of quantum fluctuations around the classical saddle point: for instance, in absence of other fields it represents the contributions of thermal gravitons on the background. If we ignore this term, we find a very interesting relation [GH77]:

$$I = \hbar \beta F. \quad (3.25)$$

This equation goes under the name of *quantum statistical relation* (the “quantum” bit refers to Planck constant), and is used in conjunction with the canonical statistical relation, which relates the free energy of the thermal system with the entropy and energy

$$F = E - TS. \quad (3.26)$$

The quantum statistical relation is a far-reaching statement that relates thermodynamics and quantum gravity. It is valid, as is the entire semiclassical approximation, if we

¹⁵But we keep $t_E = \beta \hbar$ fixed (see (3.3)). There are two other limit one can take by instead considering $\beta \hbar \rightarrow 0$: the *classical limit* ($\hbar \rightarrow 0$, β fixed) and the *high temperature limit* ($\beta \rightarrow 0$, \hbar fixed). In both these cases, the weight in the path integral reduces to the contribution of the potential evaluated on the paths.

can ignore the higher order terms in the expansion of the action, which requires the “gravitational coupling” is small, or, in terms of the Planck energy introduced at the beginning of the notes, that the energies involved are much smaller than the Planck energy. Equivalently, that the length scales of the problem are much larger than the Planck length

$$E \ll E_P \quad \Leftrightarrow \quad \ell \gg \ell_P = \sqrt{\frac{G\hbar}{c^3}}. \quad (3.27)$$

Even though it may seem that the path integral reformulation has not provided us with much insight into the quantum structure of gravity, we should not dismiss the importance of the conceptual framework that it provides us, as it will give us a lot of mileage.

3.3 Hawking–Page transition

We focus on the quantum description of a static thermal system with temperature $T = 1/\beta$ in pure gravity. Upon Wick rotation, “time” becomes a compact direction S^1 with length β , and in order to follow the path integral prescriptions we should choose a boundary ∂M , which we fix to have topology $S^1 \times S^2$. Therefore, the problem has been mapped to the question of finding metrics filling $S^1_\beta \times S^2$. In the semiclassical approximation, these metrics should also be solutions of the classical equations of motion coming from the action

$$S_E = -\frac{1}{16\pi} \int_M (R - 2\Lambda) \sqrt{g_E} d^4x - \frac{1}{8\pi} \int_{\partial M} K \sqrt{h_E} d^3x, \quad (3.28)$$

namely they are *Einstein manifolds* satisfying

$$R_{ab} = \Lambda g_{ab}. \quad (3.29)$$

We are then interested in computing the on-shell action on these solutions.

If we focus on asymptotically flat spacetime ($\Lambda = 0$), then the Schwarzschild solution will be the dominant contribution (in absence of angular momentum), but, as we mentioned in section 2.5, we find an issue with its interpretation as an equilibrium solution, because the black hole evaporates. One can also compute the specific heat of the solution, and confirm that the canonical ensemble is unstable, as the Schwarzschild black hole cannot be in thermal equilibrium with a reservoir. The temperature is given

by the inverse of β in (2.42), from which

$$T = \frac{1}{8\pi M} \quad \Rightarrow \quad C = \frac{\partial E}{\partial T} = -\frac{1}{8\pi T^2}, \quad (3.30)$$

which is negative.

A much better setup is found by looking at asymptotically *AdS* solutions ($\Lambda < 0$). As we will see momentarily, black holes in *AdS* are “eternal,” the specific heat can be positive, so they can reach equilibrium with their own thermal Hawking radiation. Intuitively, this is due to the negative gravitational constant, which acts like a “confining box” that prevents bulk objects from reaching “infinity.” We therefore focus on the study of gravity in presence of a negative cosmological constant, which we normalize to $\Lambda = -3/\ell^2$, so that

$$\begin{aligned} S_E &= -\frac{1}{16\pi} \int_M \left(R + \frac{6}{\ell^2} \right) \sqrt{g_E} d^4x - \frac{1}{8\pi} \int_{\partial M} K \sqrt{h_E} d^3x, \\ R_{ab} &= -\frac{3}{\ell^2} g_{ab}. \end{aligned} \quad (3.31)$$

3.3.1 Thermal AdS

The first “obvious” solution to (3.31) is (*Euclidean*) *anti-de Sitter spacetime*, also called *hyperbolic space*, which is the “basic” space of negative constant curvature, that is, its Riemann tensor satisfies

$$R_{abcd} = -\frac{1}{\ell^2} (g_{ac}g_{bd} - g_{ad}g_{bc}), \quad (3.32)$$

where the constant ℓ is referred to as *radius*, by analogy with the sphere (which satisfies the same condition with a positive sign). Any other constant curvature space with negative curvature is locally isometric to *EAdS*₄. The topology of *EAdS*₄ is the trivial one, same as \mathbb{R}^4 , and one can construct a coordinate system such that the metric is

$$ds^2 = \left(1 + \frac{r^2}{\ell^2} \right) dt_E^2 + \frac{dr^2}{1 + \frac{r^2}{\ell^2}} + r^2 d\Omega_2^2, \quad (3.33)$$

with $t_E \in \mathbb{R}$ and $r \geq 0$. In these coordinates, the metric on *AdS*₄ has the same form as (2.38), with the difference that here

$$f(r) = 1 + \frac{r^2}{\ell^2} \quad (3.34)$$

never vanishes, so ∂_{t_E} is never zero, and therefore there is no regularity requirement that would impose a specific periodicity for t_E . Therefore, in order to match our boundary conditions, we can identify $t_E \sim t_E + \beta$ with any β . The resulting space is referred to as *thermal AdS*.

Following the prescriptions of the gravitational path integral, we consider a region M of spacetime defined by $\{0 \leq r \leq r_0\}$ and bounded by the hypersurface $\partial M = \{r = r_0\}$ with induced metric

$$h_E = f(r_0) dt_E^2 + r_0^2 d\Omega_2^2. \quad (3.35)$$

Because of the periodic identification of t_E , this surface has the required topology $S^1 \times S^2$: the length of the circle is $\sqrt{f(r_0)}\beta$ and the radius of the 2-sphere is r_0 . As $r_0/\ell \rightarrow \infty$ and we look at “infinity,” both the circle and the sphere grow without bound:

$$h_E \sim \frac{r_0^2}{\ell^2} [dt_E^2 + \ell^2 d\Omega_2^2], \quad (3.36)$$

which is *conformally related* to a metric with finite volume $S_\beta^1 \times S_\ell^2$. This is a feature of spaces that asymptotically have the same behaviour as anti-de Sitter (unsurprisingly called asymptotically AdS spaces): we really look at their *conformal* boundary. We will return to this in section [\[\]](#). Note that when the cosmological constant vanishes, instead, the radius of the circle decouples from the radius of the sphere, and we have a typical asymptotically flat behaviour.

We are interested in the on-shell value of the Einstein–Hilbert action with Gibbons–Hawking–York term [\(3.31\)](#) The bulk term on-shell reduces to a multiple of the volume of M

$$I_{\text{bulk}} = \frac{1}{16\pi} \int_M \frac{6}{\ell^2} \sqrt{g_E} d^4x = \frac{\beta}{2\ell^2} r_0^3. \quad (3.37)$$

To compute the Gibbons–Hawking–York term we can borrow the result from [\(3.16\)](#)

$$\begin{aligned} I_{\text{GHY}} &= -\frac{1}{8\pi} \int_{\partial M} K \sqrt{h_E} d^3x = -\frac{\beta}{2} \left(\frac{r_0^2}{2} f'(r_0) + 2r_0 f'(r_0) \right) \\ &= -\frac{\beta}{2\ell^2} (3r_0^3 + 2r_0\ell^2), \end{aligned} \quad (3.38)$$

and the naive result for the on-shell action is

$$I = \frac{\beta}{2G\ell^2} (-2r_0^3 - 2r_0\ell^2). \quad (3.39)$$

This expression is divergent if $r_0 \rightarrow \infty$, corresponding to the fact that we are integrating on a non-compact space.

A surprising feature of asymptotically anti-de Sitter spaces is that the divergences that arise when computing $I_{\text{bulk}} + I_{\text{GHY}}$ can *always* be cancelled by counterterms computed using only the *intrinsic* geometry of ∂M . That is, they are integrals of the induced metric h_E , the curvature and its derivative. The resulting expressions are universal, in the sense that they apply to any asymptotically AdS solution in a given dimension.

The computation of the counterterms is a technique called *holographic renormalization*, which is always quite technical and sometimes quite subtle (see [dHSS00] for some clear explanations). The result relevant for us is that the on-shell action of *any* asymptotically (locally) $EAdS_4$ solution is finite if we add to $I_{\text{bulk}} + I_{\text{GHY}}$ the counterterm action

$$I_{\text{ct}} = \frac{1}{8\pi} \int_{\partial M_\epsilon} \left(\frac{2}{\ell} + \frac{\ell}{2} R \right) \sqrt{h_E} d^3x, \quad (3.40)$$

and then take the limit $r_0 \rightarrow \infty$.

In order to apply this to our spacetime, we need to know the Ricci scalar of the induced metric h_E , but that's a product metric on round $S^1 \times S^2$, so the Ricci scalar is just the sum of the two and one of them vanishes (being a one-dimensional metric)

$$R = \frac{2}{r_0^2}, \quad (3.41)$$

and

$$\begin{aligned} I_{\text{ct}} &= \frac{\beta}{2} \left(\frac{2}{\ell} + \frac{\ell}{r_0^2} \right) r_0^2 \sqrt{f(r_0)} \\ &= \frac{\beta}{2\ell^2} [2r_0^3 + 2r_0\ell^2 + o(1)] . \end{aligned} \quad (3.42)$$

It's clear that we cancel the divergent terms in (3.39), and in fact we just find that the on-shell action of four-dimensional thermal AdS vanishes:

$$I_{\text{Th.AdS}} = 0. \quad (3.43)$$

Quite remarkably, this is only a property of the particular choice of ∂M ! The action of $EAdS$ in a different “slicing” may be non-zero. For instance, the action of $EAdS_4$ with S^3 boundary is

$$I_{\text{EAdS}, S^3} = \frac{\pi\ell^2}{2}. \quad (3.44)$$

3.3.2 Static black hole

There are other static solutions filling the same (conformal) boundary $S^1_\beta \times S^2$. A particularly important one is the Euclidean AdS-Schwarzschild solution, which has line element

$$ds^2 = f(r) dt_E^2 + \frac{dr^2}{f(r)} + r^2 d\Omega_2^2, \quad f(r) = 1 - \frac{2M}{r} + \frac{r^2}{\ell^2}, \quad (3.45)$$

which is again a static spherically symmetric metric of the form (2.38). As before, we consider a region M of the space that has a cutoff at $r = r_0$ with induced metric that has again the same form as (3.35). Importantly, though $f(r_0)$ is different, it has the same asymptotic behaviour as (3.36) when $r_0/\ell \gg 1$: the space is asymptotically AdS, and near “infinity” it has a conformal boundary $S^1_\beta \times S^2_\ell$.

In order to have a good contribution to the semi-classical approximation of the Euclidean gravity path integral, we should also make sure that it’s a smooth spacetime. It is not obvious that the Euclidean metric is regular everywhere, because $f(r)$ now has zeroes. Concretely, we denote by r_+ the largest of the real roots of $f(r)$, which satisfies

$$M = \frac{r_+}{2\ell^2}(r_+^2 + \ell^2), \quad (3.46)$$

and then we should restrict $r \geq r_+$. In contrast to the case of thermal AdS , now our regularity discussion really parallels that around (2.38): the space is smooth if and only if we impose a periodic identification of t_E as in (2.42):

$$t_E \sim t_E + \beta, \quad \beta = \frac{4\pi r_+ \ell^2}{\ell^2 + 3r_+^2}. \quad (3.47)$$

The resulting geometry is the product of a disc and a 2-sphere described below (2.42) and represented in figure 4 (with the provision that the asymptotic behaviour is different: the radius of the circle parametrized by t_E grows rather than remaining constant, so we don’t have a “cigar.”

From the Lorentzian viewpoint, the spacetime described by (3.45) after the analytic continuation $t = -it_E$ is a static spherically symmetric black hole. There is a singularity at $r = 0$, as can be checked computing the Kretschmann scalar

$$R_{abcd}R^{abcd} = \frac{48M^2}{r^6} + \frac{24}{\ell^2},$$

and, as we showed in section 2.4, there is a Killing horizon $\{r = r_+\}$ for $k = \partial_t$, with temperature

$$T = \frac{f'(r_+)}{4\pi} = \frac{1}{4\pi r_+} + \frac{3r_+}{4\pi\ell^2} . \quad (3.48)$$

The entropy of the horizon is still given by the Bekenstein–Hawking formula for the area of the S^2 bifurcation surface (in Lorentzian), or, equivalently, the S^2 over the origin of the disc (in Euclidean)

$$S_{\text{BH}} = \frac{1}{4} A_{S^2_{r_+}} = \pi r_+^2 , \quad (3.49)$$

though we have only “rigorously” derived this formula the context of asymptotically flat solutions.

The next step is the computation of the on-shell action. At the formal level, we can use many of the expressions already derived for thermal AdS , though with different $f(r)$ and different range of r . The bulk on-shell contribution is not (3.37) because r does not extend to 0

$$I_{\text{bulk}} = \frac{\beta}{2\ell^2} (r_0^3 - r_+^3) , \quad (3.50)$$

and the Gibbons–Hawking–York term is formally equal to the first line in (3.38)

$$\begin{aligned} I_{\text{GHY}} &= -\frac{\beta}{2} \left(\frac{r_0^2}{2} f'(r_0) + 2r_0 f'(r_0) \right) \\ &= -\frac{\beta}{2\ell^2} (3r_0^3 + 2r_0\ell^2 - 3M\ell^2) . \end{aligned} \quad (3.51)$$

Overall, the divergent part is the same as that of thermal AdS , with the addition of a finite term

$$I = \frac{\beta}{2\ell^2} (-2r_0^3 - 2r_0\ell^2 + 3M\ell^2 - r_+^3) . \quad (3.52)$$

To remove the divergences we once again consider the counterterms found via holographic renormalization (3.40), which in the static spherically symmetric background have the structure in the first line of (3.42)

$$\begin{aligned} I_{\text{ct}} &= \frac{\beta}{2} \left(\frac{2}{\ell} + \frac{\ell}{r_0^2} \right) r_0^2 \sqrt{f(r_0)} \\ &= \frac{\beta}{2\ell^2} [2r_0^3 + 2r_0\ell^2 - 2M\ell^2 + o(1)] . \end{aligned} \quad (3.53)$$

Finally, taking the limit $r_0 \rightarrow \infty$, we find

$$I_{\text{AdS-Schw}} = \frac{\beta}{2} \left(M - \frac{r_+^3}{\ell^2} \right) = \frac{\beta}{4\ell^2} r_+ (\ell^2 - r_+^2) . \quad (3.54)$$

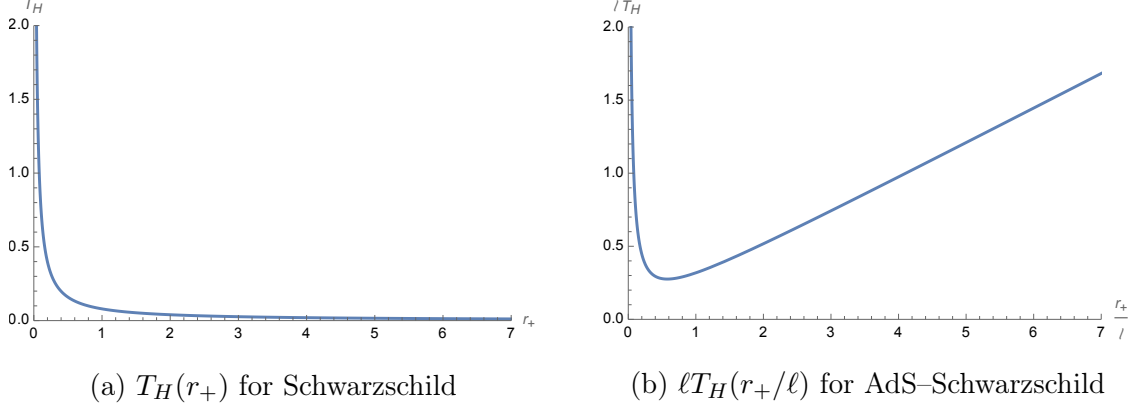


Figure 7: Plots of temperature against the horizon radius for static spherical black holes in asymptotically flat and asymptotically AdS spacetime.

This is where the physics gets more interesting! We aim to describe the canonical ensemble defined by the gravitational path integral with boundary conditions fixing the temperature, and we have found two competing leading contributions. We should therefore express the parameter of the solution M in terms of T . In fact, we have already traded M for r_+ via (3.46), so we need to invert $T = T(r_+)$ in (3.48). The expression for the temperature now has a competition between the two scales of the problem (r_+, ℓ) that is absent in Schwarzschild (which can be obtained taking $\ell \rightarrow \infty$, thus removing the cosmological constant). The result is that there is a minimum temperature at

$$r_{\min} = \frac{\ell}{\sqrt{3}}, \quad T_{\min} = \frac{\sqrt{3}}{2\pi\ell}. \quad (3.55)$$

Looking at the plot 7b, we see that for $T > T_{\min}$ there are two black hole solutions. Since one branch corresponds to a smaller radius of the other branch, the two solutions are referred to as *small* and *large* black holes. When inverting $T(r_+)$ we find that they correspond to the two roots

$$r_+^{L,S} = \frac{\ell}{3} \left(2\pi\ell T \pm \sqrt{4\pi^2\ell^2 T^2 - 3} \right), \quad (3.56)$$

which are real indeed only if $T > T_{\min}$. This equation, together with the relation $M(r_+)$ in (3.46), allows us to write $I(\beta)$ for both branches of solutions

$$I^{L,S}(\beta) = \frac{\ell \left(-\pi\ell(8\pi^2\ell^2 - 9\beta^2) \mp (4\pi^2\ell^2 - 3\beta^2)^{\frac{3}{2}} \right)}{27\beta^2}, \quad (3.57)$$

where β ranges only in $0 < \beta \leq 2\pi\ell/\sqrt{3}$. We then find that

$$I^S(\beta) > 0 \Leftrightarrow 0 < \beta \leq \frac{2\pi\ell}{\sqrt{3}}, \quad I^L(\beta) > 0 \Leftrightarrow \pi\ell < \beta \leq \frac{2\pi\ell}{\sqrt{3}}. \quad (3.58)$$

The same conclusion can also be drawn less directly from the expression (3.54). It's clear that the on-shell action of the black hole is negative if $r_+ > \ell \equiv r_{\text{HP}}$, corresponding to the temperature

$$T_{\text{HP}} = \frac{1}{\pi\ell}. \quad (3.59)$$

One then checks for consistency that $T_{\text{HP}} > T_{\text{min}}$ and that $r_{\text{HP}} > r_{\text{min}}$, which means that we are necessarily looking at the branch of large black holes.

We are now ready to discuss the thermodynamics of the system using the quantum statistical relation (3.25). We first observe that for $T < T_{\text{min}}$ there is a unique solution with the appropriate boundary condition, thermal *AdS*, so the only possible equilibrium is without a black hole. As T is raised above T_{min} there are two additional possible solution that compete with thermal *AdS* as equilibria. When $T_{\text{min}} < T < T_{\text{HP}}$, the free energies of both small and large black holes are positive, so it is still most favorable that the black hole evaporates leaving only thermal *AdS*. Finally, if $T > T_{\text{HP}}$ the free energy of the large black holes becomes lower than that of thermal *AdS* and the stable thermodynamic equilibrium is the large black hole state. This is represented in the diagram 8.

Since the free energy is continuous but its first derivative is not, this change is a first order phase transition, the *Hawking–Page transition* [HP83]. A remarkable observation from the viewpoint of the gravitational path integral is that the phase transition also involves a transition between solutions with different topologies! If the length of the boundary circle is larger than $\beta_{\text{HP}} = \pi\ell$, the dominant contribution to the path integral is the one “filling the sphere” with topology $S^1 \times \mathbb{R}^3$, as $\beta < \beta_{\text{HP}}$, instead, the dominant contribution is that “filling the circle” with topology $\mathbb{R}^2 \times S^2$.

To confirm the previous observations we should also establish the stability of the equilibrium configurations, which is done by considering the heat capacity

$$C = \frac{\partial \langle E \rangle}{\partial T}, \quad (3.60)$$

where

$$\langle E \rangle = -\frac{\partial}{\partial \beta} \log Z(\beta). \quad (3.61)$$

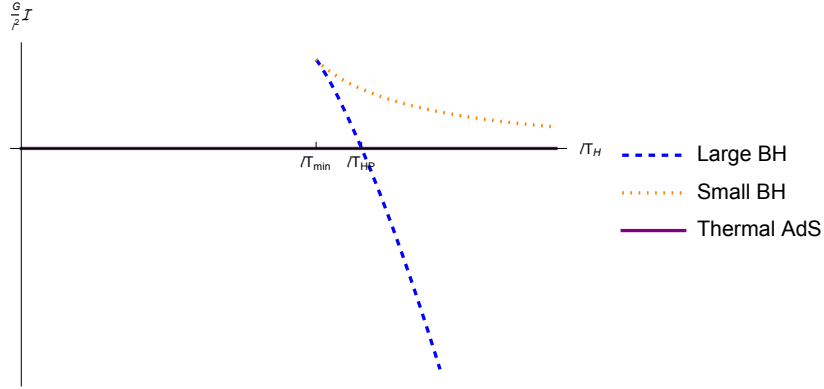


Figure 8: On-shell action of the three competing solutions with asymptotic boundary $S^1 \times S^2$ as a function of the temperature. Large and small black holes only exist if $T > T_{\min}$, and for $T > T_{\text{HP}}$ the large black hole is thermodynamically favorable compared to thermal AdS .

One could directly apply the derivatives on the expressions for $\log Z(\beta) = -I(\beta)$ in (3.57), but this leads to cumbersome expressions. On the other hand, a quicker way is to trade β for r_+ using (3.48), and write the action (3.54) in terms of r_+

$$I_{\text{AdS-Schw}} = \pi r_+^2 \frac{\ell^2 - r_+^2}{\ell^2 - 3r_+^2}. \quad (3.62)$$

We then find that

$$\langle E \rangle = M, \quad S = \left(-1 + \beta \frac{\partial}{\partial \beta} \right) I = \pi r_+^2. \quad (3.63)$$

The first relation gives us the physical interpretation of the parameter M of the solution, and the second one confirms that the thermodynamical entropy agrees with the Bekenstein–Hawking formula (3.49) even for asymptotically AdS solutions. This is a non-trivial consistency check of the gravitational path integral approach. Finally, we compute the heat capacity

$$C = 2\pi r_+^2 \frac{3r_+^2 + \ell^2}{3r_+^2 - \ell^2}, \quad (3.64)$$

which is positive provided $r_+ > \ell/\sqrt{3}$. This is exactly r_{\min} in (3.55), so if the black hole has $r_+ > r_{\min}$, that is, it's a large black hole, then its heat capacity is positive and the canonical ensemble is well-defined because it can be in equilibrium with a reservoir held at finite temperature, unlike what happens with small black holes in AdS and with the Schwarzschild solution.

4 Selected topics on the gravitational path integral

In the previous section, we introduced the gravitational path integral as a framework to quantize gravity, and we saw that even the crudest approximation, just keeping the leading saddle in the semiclassical approximation in (3.23), led to a consistent picture for both the temperature and entropy of black holes. Moreover, we saw interesting physics emerging from the competition between saddles with different topology in anti-de Sitter.

However, not everything is smooth sailing with the gravitational path integral defined as we did: there are numerous issues that we still don't understand in full, and are the subject of current research. In this section, we review some of them.

4.1 Entropy from topology

4.2 Rotation, charge and complex metric

Until now we have only seen Wick-rotated Lorentzian spacetime with vanishing mixed terms g_{tx^i} , and so after the analytic continuation $t = -it_E$ the resulting metric was positive-definite (commonly referred to as “Euclidean” or, more properly, Riemannian). We now introduce rotation and charge, and we'll see some changes.

In Lorentzian signature, the presence of angular momentum is described in terms of a spacelike Killing vector m^a generating a $U(1)$ action on the space. If we impose that the metric near the boundary is $S^1_\beta \times_f S^2$ (where the underscore f signals the possibility of a cross term), then the $U(1)$ generated by m^a could be the azimuthal rotations of the S^2 factor. We can combine m^a with the timelike Killing vector k^a to define $\xi^a = k^a + \omega m^a$. Then, ω can be interpreted as the angular velocity: we construct adapted coordinates $k = \partial_t$ and $m = \partial_\phi$, then along an orbit of ξ^a , up to a choice of integration constant, we have $\phi = \omega t$, so we identify ω as an angular velocity.

To discuss charge, instead, we should impose gauge-invariant boundary conditions for the electromagnetic field. In four dimensions, assuming a boundary with topology $S^1_\beta \times S^2$, the gauge field is specified by imposing the flux of the curvature through S^2 (corresponding to the magnetic charge, which we set to zero), and the holonomy of the gauge potential around S^1 : $\exp(i \int_{S^1} A) \equiv \exp(i\beta\Phi)$, and we will identify Φ as the electrostatic potential. Notice that $\Phi \in \mathbb{R}$ if and only if A_{t_E} is real too, which means that the gauge field in Lorentzian signature is complex. Conversely, one could start

from a real Lorentzian gauge field with $A_t \in \mathbb{R}$, but then the Wick-rotated gauge field would be complex!

We now show that introducing rotation means that the periodic identification of the fields in the KMS condition (2.14) are modified [GP78]. To do so, we consider the grand canonical Gibbs' formula with two operators generating $U(1)$ symmetries: the charge and angular momentum (which generalizes (2.9))

$$\langle \mathcal{O} \rangle_{\beta, \Phi, \omega} \equiv Z^{-1} \text{Tr} \left(\mathcal{O} e^{-\beta(H - \Phi Q - \omega J)} \right), \quad (4.1)$$

and consider the Wightman functions of a complex scalar field $\varphi(t, \phi, \mathbf{x})$ (where \mathbf{x} are the other spatial coordinates, say r, θ) on which

$$\begin{aligned} e^{i\alpha Q} \varphi(t, \phi, \mathbf{x}) e^{-i\alpha Q} &= e^{i\alpha q} \varphi(t, \phi, \mathbf{x}), \\ e^{i\alpha J} \varphi(t, \phi, \mathbf{x}) e^{-i\alpha J} &= \varphi(t, \phi - \alpha, \mathbf{x}). \end{aligned} \quad (4.2)$$

Note that the second one is a spacetime $U(1)$ symmetry, in contrast to the first one. Then, using the Heisenberg evolution of the fields and the cyclicity of the trace, we can show that

$$\begin{aligned} G_+^{\beta, \Phi, \Omega}(t, \phi, \mathbf{x}, \mathbf{y}) &\equiv Z^{-1} \text{Tr} \left[\varphi(t, \phi, \mathbf{x}) \varphi^\dagger(0, 0, \mathbf{y}) e^{-\beta(H - \Phi Q - \omega J)} \right] \\ &= e^{\beta \Phi q} Z^{-1} \text{Tr} \left[\varphi^\dagger(0, 0, \mathbf{y}) \varphi(t + i\beta, \phi + i\beta\omega, \mathbf{x}) e^{-\beta(H - \Phi Q - \omega J)} \right] \\ &= G_-^{\beta, \Phi, \omega}(t + i\beta, \phi + i\beta\omega, \mathbf{x}, \mathbf{y}), \end{aligned} \quad (4.3)$$

which is a generalization of (2.14). In terms of the Riemannian geometry at the Wick-rotated boundary, we should be imposing a *twisted* identification of the coordinates on the boundary $\partial M = S^1 \times S^2$ with metric given by the product of the two round metrics:

$$(t_E, \phi) \sim (t_E, \phi + 2\pi) \sim (t_E + \beta, \phi - \beta\Omega). \quad (4.4)$$

Compared to the case of the static charged black hole, we specify an additional “chemical potential” $\Omega = -i\omega$ at the boundary, corresponding to a $U(1)$ symmetry of the solution. We are thus in a grand canonical ensemble and we are looking for a free energy $\beta F(\beta, \Omega, \Phi) = -\log Z(\beta, \Omega, \Phi)$. However, we now have an angle coordinate on S^2 that should be identified modulo a pure imaginary periodicity, if the Lorentzian angular velocity $\omega \in \mathbb{R}$.

Concretely, the paradigmatic spacetime describing a rotating electrically charged object in flat space is the Kerr–Newman solution, which has the line element

$$\begin{aligned} ds^2 = & -\frac{\Delta - a^2 \sin^2 \theta}{\Sigma} dt^2 - 2a \sin^2 \theta \frac{r^2 + a^2 - \Delta}{\Sigma} dt d\phi \\ & + \frac{(r^2 + a^2)^2 - \Delta a^2 \sin^2 \theta}{\Sigma} \sin^2 \theta d\phi^2 + \frac{\Sigma}{\Delta} dr^2 + \Sigma d\theta^2, \end{aligned} \quad (4.5)$$

and the gauge field

$$A = -\frac{Qr}{\Sigma} (dt - a \sin^2 \theta d\phi) + \gamma dt. \quad (4.6)$$

Here γ is a constant gauge transformation, and

$$\Sigma \equiv r^2 + a^2 \cos^2 \theta, \quad \Delta \equiv r^2 - 2Mr + a^2 + Q^2. \quad (4.7)$$

This metric and gauge field are real if the parameters (M, Q, a) are real, which means that upon Wick rotation, this metric will not be Riemannian but complex, since the terms $g_{t_E\phi}$ will be pure imaginary, and the gauge field as well. It is sometimes called a “quasi-Euclidean” metric.¹⁶ How can we analyse this case? What is the regularity that one should impose on a complex solution?

One way out of this impasse is to analytically continue both a and Q : one defines $a = -i\alpha$, $Q = -iq$, thus obtaining a Riemannian metric and a real gauge field [GH79]. We pick as ∂M a surface with constant $r = r_0$, and consider the induced metric and gauge field. As $r_0 \rightarrow \infty$, they simplify and we find

$$h_E = dt_E^2 + r^2 (d\theta^2 + \sin^2 \theta d\phi^2), \quad A = i\gamma dt_E, \quad (4.8)$$

and imposing the identifications (4.4) we match the required boundary conditions, provided $i\gamma = \Phi$. Following now the general prescription found earlier, we expect to describe the thermodynamics of the system imposing that the classical solution is smooth. After the analysis, we would analytically continue the parameters back to their Lorentzian values.

Imposing smoothness of the metric means that the topology is locally the product of a cigar and a 2-sphere. The Killing vector

$$\xi_E = \frac{\partial}{\partial t_E} - \Omega \frac{\partial}{\partial \phi} \quad (4.9)$$

¹⁶Formally, this is a class of metrics that when put in the ADM form have pure imaginary shift vectors N^i (see (4.24)).

generates a $U(1)$ isometry of the spacetime: its orbits are circles in the cigar factor that shrink as we go to $r = r_+$, since $|\xi_E|^2|_{r=r_+} = 0$, and thus ξ_E vanishes there: the fixed point set is the S^2 in the transverse space. This corresponds to the fact that in the Lorentzian metric (4.5), the locus $\{r = r_+ \equiv M + \sqrt{M^2 - a^2 - Q^2}\}$ is an event horizon for the spacetime, and it is generated by the Killing vector

$$\xi = \frac{\partial}{\partial t} + \frac{a}{r_+^2 + a^2} \frac{\partial}{\partial \phi} = i\xi_E|_{t_E=it, \alpha=ia, q=iQ} \ , \quad (4.10)$$

Introducing the thermodynamical potentials via the identifications (4.4), together with smoothness of the gauge field at the origin, gives

$$\beta = 4\pi \frac{r_+^2 - \alpha^2}{\Delta'(r_+)} = 2\pi \frac{r_+^2 - \alpha^2}{(r_+ - M)} \ , \quad \Omega = \frac{\alpha}{r_+^2 - \alpha^2} \ , \quad \Phi = -\frac{qr_+}{r_+^2 - \alpha^2} \ . \quad (4.11)$$

These formulae should be viewed as fixing the parameters of the solution (M, q, α) in terms of the boundary data (β, Φ, Ω) . So, in principle one should invert them when working (as we are) in the grand canonical ensemble and express the on-shell action in terms of the potentials (β, Φ, Ω) . This is impossible to do analytically. However, one can compute the on-shell action subtracting the value of flat space, and check that the final result

$$I = \frac{\beta}{2G} (M - q\Phi) \quad (4.12)$$

satisfies the quantum statistical relation (3.25) in the form

$$I(\beta, \Phi, \Omega) = -S_{\text{BH}} + \beta \langle E \rangle - \Phi \langle Q_e \rangle - \Omega \langle J \rangle \ , \quad (4.13)$$

where

$$\langle E \rangle = \frac{M}{G} \ , \quad \langle Q_e \rangle = \frac{q}{G} \ , \quad \langle J \rangle = -\frac{\alpha M}{G} \ , \quad (4.14)$$

and S_{BH} is the Bekenstein–Hawking entropy given by the area of the (not round) two-sphere $S_{r_+}^2$ where ξ_E in (4.9) vanishes

$$S_{\text{BH}} = \frac{\pi}{G} (r_+^2 - \alpha^2) \quad (4.15)$$

Moreover, one can check that the variables $(\langle E \rangle, \langle Q_e \rangle, \langle J \rangle)$ are indeed canonically conjugate to the thermodynamic potentials (β, Φ, Ω) . This means that can now view (4.13) as a Legendre transform from the grand canonical to the microcanonical ensemble, and

read off the thermodynamic entropy, which (again) coincides with the Bekenstein–Hawking formula (4.15), naively confirming once again the general prescription of the gravitational path integral.

Despite the apparent success of the computations, we have bypassed the subtleties of the complex metric by performing analytic continuation of the parameters, but the interpretation of the ensemble is now quite obfuscated. The problem we started with was the description of the grand canonical ensemble in the real Lorentzian background (4.5), that is, the matrix element of the (un-normalized) density matrix

$$\rho = e^{-\beta(E - \Phi Q_e - \Omega J)} . \quad (4.16)$$

The computations done with the Riemannian metric resulted in the relations (4.11) between the boundary conditions and the parameters, which in terms of the original constants in the Lorentzian metric become pure imaginary

$$\beta = 2\pi \frac{r_+^2 + a^2}{r_+ - M} , \quad \Phi = -i \frac{Q r_+}{r_+^2 + a^2} , \quad \Omega = i \frac{a}{r_+^2 + a^2} . \quad (4.17)$$

Therefore, the Riemannian metric actually corresponds to a saddle of a putative grand canonical ensemble based on the (un-normalized) density matrix

$$\rho = e^{-\beta(E - i|\Phi|Q_e - i|\Omega|J)} . \quad (4.18)$$

Not only is the physical meaning of this operator not clear, but even its mathematical properties are not obvious, since evaluated on the states it would correspond to complex weights in Gibbs’ formula (2.9), so it would not be clearly convergent.

An alternative approach is to include in the path integral the complex “quasi-Euclidean” metric obtained performing the Wick rotation $t = -it_E$ but without analytic continuation of the constants a and Q . However, once we open the Pandora’s box of complex metrics in the path integral, we need to define the rules for their inclusion. For instance, one could require that the density matrix (4.16) does have a convergent trace, which would impose a condition on the eigenvalues of $E - \Phi Q_e - \Omega J$. As it turns out, in flat space such condition is generically violated, as one can make J arbitrarily large by going sufficiently far, and the quasi-Euclidean Kerr–Newman metric is not a good saddle of the path integral.

4.3 Subtleties

We delved into the “allowability” of complex metrics in the path integral due to rotation, but there are more subtleties in the Euclidean quantum gravity approach, which we have until now steered clear of and are the subject of active research. We’ll mention here a few without attempting to propound any solution.

4.3.1 Inclusion of multiple topologies

In the discussion of quantum field theory on Lorentzian curved spacetime, we usually restrict ourselves to globally hyperbolic spacetime. These spacetimes have a Cauchy surface Σ such that the domain of dependence (which, informally, is the union of the points that are reached from Σ via a causal curve) is the entire manifold. Examples of globally hyperbolic manifolds are flat space and the Kruskal extension of Schwarzschild, but many more are known. Global hyperbolicity is our notion of “good behaviour” for a spacetime, since given data on Σ we can compute the solution to hyperbolic equations everywhere on M . Moreover, it excludes obviously problematic situations, such as closed timelike curves. Globally hyperbolic spacetimes are “good” for various reasons, among which is the existence of a global time function such that surfaces of constant t are Cauchy surfaces with topology Σ and the topology of spacetime is $\mathbb{R}_t \times \Sigma$. After Wick rotation, this leads us naturally to consider Euclidean solutions with (asymptotic) boundary $S^1_{t_E} \times \partial\Sigma$, which encode thermal properties of the Lorentzian system. More generally, the $S^1_{t_E}$ factor intuitively represents the existence in Lorentzian of a trace over a Hilbert space defined on the spatial slice Σ .

However, there are plenty of *gravitational instantons*: Riemannian regular solutions with finite Einstein–Hilbert action (with GHY term and regulator) and (asymptotic) boundary with topology different from $S^1 \times \partial\Sigma$.¹⁷ For instance, we could have smooth solutions which interpolate between different topologies (think about pair of pants) that would be singular in Lorentzian, or even smooth solutions with boundaries that do not at all have a well-defined Lorentzian continuation (for instance, they could have

¹⁷Their name comes from the analogy with *instantons* in quantum field theory (and quantum mechanics), where the name denotes solutions to the Euclidean equations of motion with finite action. Schematically, these represent tunneling effects between different vacua of the theory.

topology $M \cong \mathbb{R}^4$ and fill a $\partial M \cong S^3$, or even more exotic three-manifolds).¹⁸ What should we make of these solutions?

4.3.2 “Conformal factor problem”

One of the reasons we put forward as an argument for the use of the Euclidean action in the path integral for quantum gravity was that in Euclidean signature the path integral of quantum field theory converges. Whilst this does hold for the theories of fields with spin 0 and 1 and for anti-commuting with spin $\frac{1}{2}$, which have positive semi-definite actions, it is not true for gravity, because the Euclidean action

$$S_E[g_E] = -\frac{1}{16\pi} \int_M R \sqrt{g_E} d^4x - \frac{1}{8\pi} \int_{\partial M} K \sqrt{h_E} d^3x \quad (4.19)$$

does not have a definite sign [GHP78]. Indeed, under a Weyl rescaling of the metric g_E to $\tilde{g}_E = \Omega^2 g_E$ (which is *not* a diffeomorphism but a change in the space of metrics), we have

$$\tilde{R} = \Omega^{-2} (R - 6\Omega^{-1} \nabla^2 \Omega) , \quad \tilde{K} = \Omega^{-1} (K + 3\Omega^{-1} n^a \nabla_a \Omega) . \quad (4.20)$$

Therefore (integrating by parts and using the divergence theorem)

$$S_E[\tilde{g}_E] = -\frac{1}{16\pi} \int_M (\Omega^2 R + 6 |\mathrm{d}\Omega|^2) \sqrt{g_E} d^4x - \frac{1}{8\pi} \int_{\partial M} \Omega^2 K \sqrt{h_E} d^3x , \quad (4.21)$$

and we can make this arbitrarily negative by choosing $|\mathrm{d}\Omega|^2$ sufficiently large, so the Euclidean gravity action is unbounded below. This problem does not arise at the leading order in the semi-classical expansion (3.23), as one is considering only a solution of the equations of motion, but it does become crucial when including the one-loop contribution.

4.3.3 Gauge group

Another problem that only arises when going beyond the leading order in (3.23) is that of gauge redundancy, because in doing the path integral one integrates over all configurations, even those that are physically equivalent because of redundancy of the description. We then use elaborate instruments to account for this (e.g. BRST or BV

¹⁸The existence of a time-orientable Lorentzian structure on a manifold requires certain topology, whereas this is not true of Riemannian metrics, which exist on any smooth manifold.

formalism), but first we must establish whether gravity has these redundancy. General relativity is indeed a gauge theory, but it is not like any gauge theory you may have seen until now, because the gauge group is the infinite-dimensional group of diffeomorphisms connected to the identity $\text{Diff}_0(M)$.¹⁹ The Lie group of diffeomorphisms is not necessarily connected: for instance if M is a Lie group G , then G acts on itself via left-translation, so $\text{Diff}(G)$ contains G , and thus may have multiple connected components. Components that are not connected to the identity are considered to act non-trivially on the theory, that is, they are not redundancies to be “gauged away.” In fact, even the action of diffeomorphisms in the component connected to the identity may be non-trivial.

Let X^a be a vector representing the infinitesimal action of a diffeomorphism in $\text{Diff}_0(M)$, and we let X^a act on the Lagrangian density $\mathcal{L}_{\text{EH}} = (R - 2\Lambda)\sqrt{g_E}$ via Lie derivative. Recalling that \mathcal{L}_{EH} is a tensor density of weight 1, we have

$$\begin{aligned}\mathcal{L}_X [(R - 2\Lambda)\sqrt{g_E}] &= X^a \partial_a (R - 2\Lambda) \sqrt{g_E} + \nabla_a X^a (R - 2\Lambda) \sqrt{g_E} \\ &= \nabla_a [X^a (R - 2\Lambda)] \sqrt{g_E},\end{aligned}\tag{4.22}$$

so we can use the divergence theorem to conclude that the variation of the Einstein–Hilbert action is a boundary term

$$\delta_X \int_M (R - 2\Lambda) \sqrt{g_E} d^4x = \int_{\partial M} n_a X^a (R - 2\Lambda) \sqrt{h_E} d^3x.\tag{4.23}$$

If the vector field does not have compact support or more generally vanishes sufficiently fast near the boundary, the action is not invariant under the diffeomorphism it generates, and so it is *not* a redundancy in the description. The diffeomorphisms that act in a non-trivial way on the theory, whether because they are not connected to the identity or because the generating vector field does not vanish at the (potentially asymptotic) boundary, are sometimes called *large* diffeomorphisms (in contrast to *small* diffeomorphisms).

This fact also emerges in the Hamiltonian formulation of general relativity. In a globally hyperbolic spacetime, as already mentioned, we have a global time function

¹⁹For the *cognoscenti*: at the end of the day, the difference with, say, Yang–Mills theory of the Lie group G , which is based on a G -principal bundle $P \rightarrow M$ is due to the fact that in GR the relevant principal bundle is the frame bundle itself.

$t : M \rightarrow \mathbb{R}$ such that $\{t = 0\}$ is a Cauchy surface, and from there we can construct additional coordinates x^i such that the metric has the form

$$ds^2 = -N^2 dt^2 + h_{ij}(dx^i + N^i dt)(dx^j + N^j dt). \quad (4.24)$$

This is referred to as “3 + 1” or ADM decomposition (Arnowitt–Deser–Misner), and shows that we can rewrite the action in terms of the *lapse function* $N(t, \mathbf{x})$, the *shift vector* $N^i(t, \mathbf{x})$ and the induced metric on the Cauchy surface at constant t , $h_{ij}(t, \mathbf{x})$. They work as “canonical positions” for the initial value problem. One then computes the canonically conjugate momenta, and it turns out that the only relevant one is the momenta canonically conjugate to h_{ij} , which is $\Pi^{ij}\sqrt{h}$, where Π^{ij} is the Brown–York tensor that already appeared in (3.19) (basically with a computation analogous to that leading to (3.18)). The resulting Hamiltonian is composed by a bulk term and a boundary term proportional to Π_{ij} , but the bulk term *vanishes* on-shell, and the only contribution comes from the boundary.

This should not come to you as a surprise. Defining energy in general relativity is not easy, because of the equivalence principle and the properties of small diffeomorphisms: in a neighbourhood of a point, you can always define normal coordinates such that any quantity defined with the metric and its first derivatives would vanish, thus “gauging away” gravity. However, as we just found out, large diffeomorphisms are not just gauge redundancies, and correspondingly the Hamiltonian does receive contributions from the boundary that can be evaluated. In fact, one can show that the variation of the action with respect to the boundary metric, that is Π_{ij} in (3.19), behaves as a stress-energy tensor (explaining its name *Brown–York stress-energy tensor*). Again, this is not a surprise: our (admittedly brief) study of black holes suggested that gravitational physics in a region should be described in terms of its boundary (the holographic principle mentioned in section 2.5), which perfectly resonating with what we just described! In fact, this is borne out in the context of the AdS/CFT correspondence: the dynamics of gravity in AdS is actually described by a theory “living” on its boundary for which the Brown–York tensor is the actual stress-energy tensor.

References

- [AHM⁺20] A. Almheiri, T. Hartman, J. Maldacena, E. Shaghoulian, and A. Tajdini, *The entropy of Hawking radiation*, [Rev. Mod. Phys. **93** \(2021\) 035002](#), [arXiv:2006.06872 \[hep-th\]](#).
- [BCH73] J. M. Bardeen, B. Carter, and S. W. Hawking, *The Four laws of black hole mechanics*, [Commun. Math. Phys. **31** \(1973\) 161–170](#).
- [BD84] N. D. Birrell and P. C. W. Davies, *Quantum Fields in Curved Space*, Cambridge Monographs on Mathematical Physics, Cambridge Univ. Press, Cambridge, UK, 2 1984.
- [Bek72] J. D. Bekenstein, *Black holes and the second law*, [Lett. Nuovo Cim. **4** \(1972\) 737–740](#).
- [Bek73] J. D. Bekenstein, *Black holes and entropy*, [Phys. Rev. D **7** \(1973\) 2333–2346](#).
- [Bou02] R. Bousso, *The Holographic principle*, [Rev. Mod. Phys. **74** \(2002\) 825–874](#), [arXiv:hep-th/0203101](#).
- [dHSS00] S. de Haro, S. N. Solodukhin, and K. Skenderis, *Holographic reconstruction of space-time and renormalization in the AdS / CFT correspondence*, [Commun. Math. Phys. **217** \(2001\) 595–622](#), [arXiv:hep-th/0002230](#).
- [DR23] I. Davies and H. S. Reall, *Nonperturbative Second Law of Black Hole Mechanics in Effective Field Theory*, [Phys. Rev. Lett. **132** \(2024\) 171402](#), [arXiv:2312.07659 \[hep-th\]](#).
- [FR87] S. A. Fulling and S. N. M. Ruijsenaars, *Temperature, periodicity and horizons*, [Physics Reports **152** \(1987\) 135–176](#).
- [GH77] G. W. Gibbons and S. W. Hawking, *Action Integrals and Partition Functions in Quantum Gravity*, [Phys. Rev. D **15** \(1977\) 2752–2756](#).
- [GH79] ———, *Classification of Gravitational Instanton Symmetries*, [Commun. Math. Phys. **66** \(1979\) 291–310](#).
- [GH93] G. W. Gibbons and S. W. Hawking (eds.), *Euclidean quantum gravity*, World Scientific, 1993.
- [GHP78] G. W. Gibbons, S. W. Hawking, and M. J. Perry, *Path Integrals and the Indefiniteness of the Gravitational Action*, [Nucl. Phys. B **138** \(1978\) 141–150](#).

- [GP78] G. W. Gibbons and M. J. Perry, *Black Holes and Thermal Green's Functions*, [Proc. Roy. Soc. Lond. A **358** \(1978\) 467–494](#).
- [Haw74] S. W. Hawking, *Black hole explosions*, [Nature **248** \(1974\) 30–31](#).
- [Haw75] ———, *Particle Creation by Black Holes*, [Commun. Math. Phys. **43** \(1975\) 199–220](#). [Erratum: Commun.Math.Phys. 46, 206 (1976)].
- [Haw78] S. W. Hawking, *Euclidean Quantum Gravity*, Recent Developments in Gravitation (M. Levy and S. Deser, eds.), Cargèse Lectures, Plenum, 1978.
- [Haw79] ———, *The Path-Integral Approach to Quantum Gravity*, General Relativity: An Einstein Centenary Survey (S. W. Hawking and W. Israel, eds.), Cambridge University Press, 1979.
- [HE23] S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space-Time*, Cambridge Monographs on Mathematical Physics, Cambridge University Press, 2 2023.
- [HP83] S. W. Hawking and D. N. Page, *Thermodynamics of Black Holes in anti-De Sitter Space*, [Commun. Math. Phys. **87** \(1983\) 577](#).
- [HW14] S. Hollands and R. M. Wald, *Quantum fields in curved spacetime*, [Phys. Rept. **574** \(2015\) 1–35](#), [arXiv:1401.2026 \[gr-qc\]](#).
- [IW94] V. Iyer and R. M. Wald, *Some properties of Noether charge and a proposal for dynamical black hole entropy*, [Phys. Rev. D **50** \(1994\) 846–864](#), [arXiv:gr-qc/9403028](#).
- [Jac03] T. Jacobson, *Introduction to quantum fields in curved space-time and the Hawking effect*, School on Quantum Gravity, 8 2003, pp. 39–89. [arXiv:gr-qc/0308048](#).
- [JKM93] T. Jacobson, G. Kang, and R. C. Myers, *On black hole entropy*, [Phys. Rev. D **49** \(1994\) 6587–6598](#), [arXiv:gr-qc/9312023](#).
- [KU22] C. Kehle and R. Unger, *Gravitational collapse to extremal black holes and the third law of black hole thermodynamics*, [arXiv:2211.15742 \[gr-qc\]](#).
- [KW91] B. S. Kay and R. M. Wald, *Theorems on the Uniqueness and Thermal Properties of Stationary, Nonsingular, Quasifree States on Space-Times with a Bifurcate Killing Horizon*, [Phys. Rept. **207** \(1991\) 49–136](#).

- [Mat09] S. D. Mathur, *The Information paradox: A Pedagogical introduction*, [Class. Quant. Grav.](#) **26** (2009) 224001, [arXiv:0909.1038 \[hep-th\]](#).
- [MTW73] C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation*, W. H. Freeman, San Francisco, 1973.
- [Pag04] D. N. Page, *Hawking radiation and black hole thermodynamics*, [New J. Phys.](#) **7** (2005) 203, [arXiv:hep-th/0409024](#).
- [Rea20] H. Reall, *Part 3 Black Holes*, [Link](#), 2020.
- [Rea24] H. S. Reall, *Third law of black hole mechanics for supersymmetric black holes and a quasilocal mass-charge inequality*, [Phys. Rev. D](#) **110** (2024) 124059, [arXiv:2410.11956 \[gr-qc\]](#).
- [Ros05] S. F. Ross, *Black hole thermodynamics*, [arXiv:hep-th/0502195](#).
- [Sch51] J. S. Schwinger, *On gauge invariance and vacuum polarization*, [Phys. Rev.](#) **82** (1951) 664–679.
- [SV96] A. Strominger and C. Vafa, *Microscopic origin of the Bekenstein-Hawking entropy*, [Phys. Lett. B](#) **379** (1996) 99–104, [arXiv:hep-th/9601029](#).
- [Tow97] P. K. Townsend, *Black holes: Lecture notes*, [arXiv:gr-qc/9707012](#).
- [Wal95] R. M. Wald, *Quantum Field Theory in Curved Space-Time and Black Hole Thermodynamics*, University of Chicago Press, 1995.
- [Wal15] A. C. Wall, *A Second Law for Higher Curvature Gravity*, [Int. J. Mod. Phys. D](#) **24** (2015) 1544014, [arXiv:1504.08040 \[gr-qc\]](#).
- [Wit21] E. Witten, *Why does quantum field theory in curved spacetime make sense? And what happens to the algebra of observables in the thermodynamic limit?*, 2022. [arXiv:2112.11614 \[hep-th\]](#).
- [Wit24] ———, *Introduction to black hole thermodynamics*, [Eur. Phys. J. Plus](#) **140** (2025) 430, [arXiv:2412.16795 \[hep-th\]](#).