

Black Holes and Semiclassical Quantum Gravity

Davide Cassani

INFN, Sezione di Padova, Via Marzolo 8, 35131 Padova, Italy

`davide.cassani@pd.infn.it`

June 23, 2021

These are lecture notes for a course on black holes and semiclassical quantum gravity, given at the LACES 2019 graduate school in Florence and at the Galilean School in Padova.

Contents

1	Motivation	1
2	Some basics	2
2.1	Conventions and Stokes' theorem	2
2.2	Komar integrals and conserved spacetime charges	4
2.3	Killing horizons and surface gravity	6
2.4	Generalized Smarr formula	10
2.5	The Kerr-Newman solution	12
2.6	Ergosphere	16
2.7	Penrose process for energy extraction	17
3	Black hole thermodynamics	19
3.1	Why should black holes carry an entropy?	19
3.2	The laws of black hole mechanics	20
4	QFT in curved spacetime and Hawking radiation	24
4.1	QFT in curved spacetime	24
4.2	Hawking radiation	27
4.3	Further remarks	32
5	Euclidean Quantum Gravity	34
5.1	QFT at finite temperature	34
5.2	Hawking temperature from regularity of Euclidean geometry	39
5.3	Regularity of Kerr-Newman and grand-canonical ensemble	40
5.4	The gravitational path integral	42
5.5	The Euclidean on-shell action	43
5.6	The on-shell action in the grand-canonical ensemble	46
6	Black holes in AdS and phase transitions	47
7	Wald's entropy	47

8	The quantum entropy of extremal black holes	48
8.1	Extremal black holes	49
8.2	The entropy function	51
8.3	Relation with Euclidean on-shell action	54
8.4	Quantum entropy	57
9	Black hole microstate counting	58
A	Statistical ensembles	61
A.1	Microcanonical ensemble	61
A.2	Canonical ensemble	61
A.3	Grand-canonical ensemble	62

References

These notes do not contain anything original, they just assemble material taken from various references, including the following ones.

Differential geometry:

- M. Nakahara, *Geometry, Topology and Physics*, CRC Press, 2003.

General relativity:

- S. Carroll, *Spacetime and Geometry. An Introduction to General Relativity*, Pearson 2013. See also the online *Lecture Notes on General Relativity*, [gr-qc/9712019](#).
- R. Wald, *General relativity*, University of Chicago Press, 1984.

More specific on Black Holes:

- P. Townsend, *Black holes: Lecture notes*, [gr-qc/9707012](#).
- H. Reall, lecture notes on *Black Holes*.
- T. Hartman, *Lectures on Quantum Gravity and Black Holes*.
- J. Polchinski, *The Black Hole Information Problem*, lectures at TASI 2015, [1609.04036](#).

More references are given throughout the notes, with the scope of pointing at the papers where the original results appeared, or to encourage further reading. Clearly, this has no purpose of completeness.

1 Motivation

Einstein's general Relativity (GR) is non-renormalizable and should be seen as the low-energy effective theory of a more fundamental, UV-complete theory of Quantum Gravity. The UV completion of GR that is realized in Nature is not known yet, string theory being a strong candidate. However, even if we don't know the UV complete theory we can ask how consistency with phenomena that we observe at low-energy puts constraints on Quantum Gravity. Conversely, any candidate fundamental theory of Quantum Gravity must be able to explain all low-energy phenomena, and we would like to test such ability. In this course, we will explore these questions using black holes.

Why are black holes relevant for Quantum Gravity? As all other solutions to the equations of general relativity, they are a priori entirely classical objects. However, a surprising feature is that they display thermodynamic properties. The laws of ordinary thermodynamics emerge as a macroscopic, coarse grained description of an ensemble of many microscopic states; using statistical mechanics, it is possible to derive these laws from the kinetic theory of gases, for instance. Similarly, the laws of black hole thermodynamics may be seen as emergent properties of gravity in the low-energy effective theory provided by GR. Understanding how black hole thermodynamics is modified as we go higher in energy may reveal us something about the fundamental theory of Quantum Gravity, thus providing a window into the quantum structure of spacetime. Conversely, it should be possible to derive the black hole thermodynamics, and the corrections to it, starting from a fundamental theory of Quantum Gravity and taking some appropriate average, or coarse-graining limit.

An important hint in this direction comes from the celebrated Bekenstein-Hawking formula expressing the black hole entropy. Including all the dimensionful constants, this formula reads

$$S = k_B \frac{A}{4\ell_P^2} = k_B \frac{c^3 A}{4\hbar G} , \quad (1.1)$$

where $\ell_P = \sqrt{\frac{\hbar G}{c^3}}$ is the Planck length and A is the area of the event horizon. This is one of the most beautiful formulae in physics, in that it brings together in a simple way quantities associated with different domains of physics: the entropy S is a thermodynamic quantity, the Boltzmann constant k_B refers to statistical physics, the Newton constant G is gravity, the speed of light c is special relativity, the Planck constant \hbar is quantum mechanics, and the area A is geometry. This seems to imply that we probably need to merge and take into account all these different subjects in order to understand the formula and derive it from a fundamental theory. In particular, the appearance of \hbar means that even if black holes are solutions of a classical theory, we need quantum mechanics to describe the microstates responsible for

their entropy. So understanding the black hole entropy is ultimately a Quantum Gravity problem.

We can be more concrete and formulate a precise question. Recall that the macroscopic entropy of a many body system with charges Q is related to the underlying microstates as

$$S(Q) = k_B \log d(Q), \quad (1.2)$$

where $d(Q)$ is the degeneracy of microstates carrying the quantum numbers Q . An important challenge for any fundamental theory of quantum gravity is to reproduce the black hole entropy formula by computing this degeneracy. This is the problem of black hole *microstate counting*.

The formula (1.1) also has another surprising feature: the black hole entropy is proportional to the area, rather than to the volume as in ordinary systems. This seems to indicate that the degrees of freedom of gravity are stored in one dimension less than in usual systems. This observation was one of the main motivations that inspired the holographic principle and eventually led to the AdS/CFT correspondence in the context of string theory. In turn, the AdS/CFT correspondence provides new tools for understanding quantum gravity. Using these tools, it has been shown in certain controlled setups that string theory provides the black hole microstates, and successfully solves the problem of microstate counting.

We will get to microstate counting only towards the end of this course. For the main part of it, we will discuss how black hole thermodynamics arises in GR, how it can be further explored using semiclassical reasoning, and what tools can be used to go beyond the semiclassical approximation and define a quantum entropy.

2 Some basics

2.1 Conventions and Stokes' theorem

- Unless otherwise specified, we take $c = \hbar = G = k_B = 1$.
- We use a mostly plus metric $(- + \dots +)$. Our convention for the Riemann curvature tensor is

$$(\nabla_\mu \nabla_\nu - \nabla_\nu \nabla_\mu)V^\rho = R_{\mu\nu}{}^\rho{}_\sigma V^\sigma, \quad (2.1)$$

which in terms of the Christoffel symbols gives

$$R_{\mu\nu}{}^\rho{}_\sigma = \partial_\mu \Gamma_{\nu\sigma}^\rho - \partial_\nu \Gamma_{\mu\sigma}^\rho + \Gamma_{\mu\lambda}^\rho \Gamma_{\nu\sigma}^\lambda - \Gamma_{\nu\lambda}^\rho \Gamma_{\mu\sigma}^\lambda. \quad (2.2)$$

The Ricci tensor is $R_{\mu\nu} = R_{\rho\mu}{}^{\rho}{}_{\nu}$, and the Ricci scalar is $R = g^{\mu\nu}R_{\mu\nu}$.

- Let us consider a d -dimensional manifold M endowed with a metric $g_{\mu\nu}$ (for most of the time we will take $d = 4$, but here we can keep the spacetime dimension general). We denote by $\epsilon_{\mu_1\dots\mu_d}$ the totally antisymmetric tensor, with $\epsilon_{0\dots d} = \sqrt{|g|}$ (so this is not the tensor *density*). It satisfies

$$\epsilon^{\mu_1\dots\mu_p\lambda_{p+1}\dots\lambda_d}\epsilon_{\nu_1\dots\nu_p\lambda_{p+1}\dots\lambda_d} = (-)^t p!(d-p)! \delta_{[\nu_1}^{\mu_1} \dots \delta_{\nu_p]}^{\mu_p}. \quad (2.3)$$

where $t = 0$ if M is Riemannian while $t = 1$ if M is Lorentzian, and the indices are raised using the inverse metric.

We denote p -forms as

$$\omega = \frac{1}{p!} \omega_{\mu_1\dots\mu_p} dx^{\mu_1} \wedge \dots \wedge dx^{\mu_p}. \quad (2.4)$$

The Hodge dual of a p -form ω on M is a $(d-p)$ -form defined as¹

$$*\omega = \frac{1}{p!(d-p)!} \omega_{\mu_1\dots\mu_p} \epsilon^{\mu_1\dots\mu_p}{}_{\mu_{p+1}\dots\mu_d} dx^{\mu_{p+1}} \wedge \dots \wedge dx^{\mu_d}. \quad (2.5)$$

The Hodge dual satisfies

$$**\omega = (-)^{t+p(d-p)} \omega, \quad (2.6)$$

where t distinguishes between a Riemannian or Lorentzian manifold as above. For $p \geq 1$, we also have

$$*d*\omega = \frac{1}{(p-1)!} (-)^{t+(p-1)(d-p)} \nabla^\nu \omega_{\nu\mu_1\dots\mu_{p-1}} dx^{\mu_1} \wedge \dots \wedge dx^{\mu_{p-1}}, \quad (2.7)$$

which expresses the divergence of a tensor in differential form language.

- **Stokes' theorem.** Given a d -dimensional manifold M with boundary ∂M and a $(d-1)$ -form ω , Stokes' theorem states that

$$\int_M d\omega = \int_{\partial M} \omega. \quad (2.8)$$

An application of this theorem is in conservation laws. Assume the spacetime is foliated by spacelike hypersurfaces Σ_t at fixed time t (Cauchy surfaces), and consider two such hypersurfaces, Σ_{t_1} and Σ_{t_2} . These bound a spacetime region M , with $\partial M = \Sigma_{t_2} \cup \Sigma_{t_1}$. Assume we have a conserved current,

$$\nabla_\mu j^\mu = 0 \quad \Leftrightarrow \quad d*j = 0, \quad (2.9)$$

¹This definition is as in Carroll, Nakahara and Wald, for instance. In other references, such as e.g. Reall's lecture notes, the $\mu_1\dots\mu_p$ and $\mu_{p+1}\dots\mu_d$ set of indices are swapped in the ϵ tensor. This leads to an opposite sign for the Hodge star of forms of odd degree in an even-dimensional spacetime.

where in the second expression $j = j_\mu dx^\mu$. The associated charge at the time t is

$$Q(t) = \int_{\Sigma_t} *j. \quad (2.10)$$

Then Stokes' theorem gives

$$0 = \int_M d*j = \int_{\partial M} *j = \int_{\Sigma_{t_2}} *j - \int_{\Sigma_{t_1}} *j \quad \Rightarrow \quad Q(t_2) = Q(t_1), \quad (2.11)$$

namely the charge is conserved. Because of this, it can be measured at any time t .

• **Electric and magnetic charges.** The Maxwell equations

$$\nabla^\nu F_{\nu\mu} = -4\pi j_\mu, \quad \nabla_{[\mu} F_{\nu\rho]} = 0 \quad (2.12)$$

read in differential form notation

$$d*j = 4\pi*j, \quad dF = 0. \quad (2.13)$$

The first implies the conservation of the current, $d*j = 0$. The second implies that locally there exists a one-form A such that $F = dA$; note that A is defined only modulo gauge transformations $A \rightarrow A + d\lambda$. Using Maxwell and then Stokes, we find

$$Q = \int_{\Sigma} *j = \frac{1}{4\pi} \int_{\Sigma} d*j = \frac{1}{4\pi} \int_{\partial\Sigma} *F. \quad (2.14)$$

This is Gauss' law in differential form language. Notice that the electromagnetic field can carry charge even in the absence of sources, namely even if $j = 0$.

We can use (2.14) to define the electric charge and magnetic charges of the whole space-time. Let us fix $d = 4$ for definiteness. Take a Cauchy surface Σ , introduce some radial coordinate r and consider the 2-sphere S_r^2 at fixed r . Then the electric charge of the space-time is defined as

$$Q = \frac{1}{4\pi} \lim_{r \rightarrow \infty} \int_{S_r^2} *F. \quad (2.15)$$

Similarly, in four dimensions we can introduce the magnetic charge P as

$$P = \frac{1}{4\pi} \lim_{r \rightarrow \infty} \int_{S_r^2} F. \quad (2.16)$$

2.2 Komar integrals and conserved spacetime charges

Let us see how to also associate conserved charges to spacetime symmetries. Here we can work in arbitrary spacetime dimension d . Assume we have a Killing vector K ; vanishing of the Lie derivative of the metric gives

$$\nabla_\mu K_\nu + \nabla_\nu K_\mu = 0. \quad (2.17)$$

It is not hard to show that²

$$\nabla_\mu \nabla_\nu K^\rho = R^\rho{}_{\nu\mu\sigma} K^\sigma. \quad (2.18)$$

Contracting the μ and ρ indices and using the Killing equation (2.17), we get

$$\nabla^\rho \nabla_\rho K_\mu = -R_{\mu\nu} K^\nu. \quad (2.19)$$

Using (2.7) to express the l.h.s. in differential form notation and using the (trace-reversed) Einstein equation $R_{\mu\nu} = 8\pi(T_{\mu\nu} - \frac{1}{d-2} g_{\mu\nu} T)$ on the r.h.s (here $T = T^\rho{}_\rho$), we arrive at

$$*d * dK = 8\pi(-)^{t+d+1} j. \quad (2.20)$$

where we have defined the one-form current

$$j_\mu = 2 \left(T_{\mu\nu} - \frac{1}{d-2} g_{\mu\nu} T \right) K^\nu. \quad (2.21)$$

It follows that j is a conserved current,

$$d * j = 0. \quad (2.22)$$

The spacetime symmetry generated by K then leads to the charge

$$Q_K = c \int_\Sigma *j = \frac{c}{8\pi} \int_\Sigma d * dK = \frac{c}{8\pi} \int_{\partial\Sigma} *dK, \quad (2.23)$$

where c is some constant. This expression is called *Komar integral*.

Recall that an *asymptotically flat* spacetime is a spacetime which looks like Minkowski space at large distance. Our working definition of asymptotic flatness is that in the coordinates t, r, θ, ϕ that we will be using, the spacetime metric looks like the one of Minkowski space, $ds^2 \sim -dt^2 + dr^2 + r^2 (d\theta^2 + \sin^2 \theta d\phi^2)$ asymptotically, namely for $r \rightarrow \infty$.

Recall that a spacetime is *stationary* if there is a Killing vector K that is everywhere timelike; in this case we can find coordinates such that $K = \partial/\partial t$. A spacetime is *axisymmetric* if it admits a spacelike Killing vector \tilde{K} generating the isometry group $U(1)$; so we can find an angular coordinate $\phi \sim \phi + 2\pi$ such that $\tilde{K} = \partial/\partial\phi$.

Consider a four-dimensional, asymptotically flat stationary spacetime. We can use the Komar integral to define the mass (or energy) by taking the integral over the spacelike sphere at infinity:

$$M_{\text{Komar}} = -\frac{1}{8\pi} \lim_{r \rightarrow \infty} \int_{S^2} *dK. \quad (2.24)$$

²To see this, in addition to the Killing equation $\nabla_\mu K_\nu + \nabla_\nu K_\mu = 0$, use $[\nabla_\rho, \nabla_\nu]K_\mu = R_{\rho\nu\mu\sigma}K^\sigma$ and $[\nabla_\rho, \nabla_\nu]K_\mu = -[\nabla_\mu, \nabla_\rho]K_\nu - [\nabla_\nu, \nabla_\mu]K_\rho$ (i.e. the first Bianchi identity of the Riemann tensor).

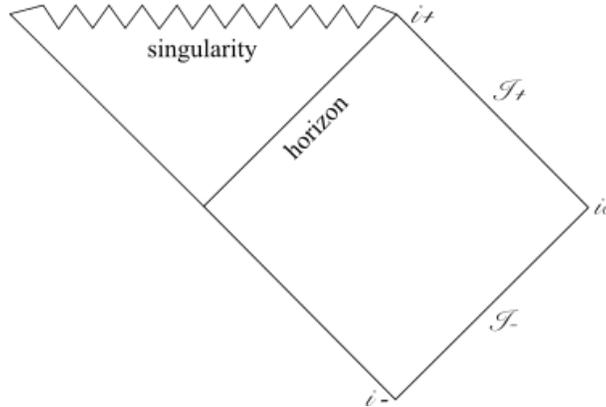
If the spacetime is also axisymmetric (with $[K, \tilde{K}] = 0$), we can define the angular momentum as

$$J_{\text{Komar}} = \frac{1}{16\pi} \lim_{r \rightarrow \infty} \int_{S_r^2} *d\tilde{K}. \quad (2.25)$$

The overall coefficients in these expressions have been fixed by taking the flat space limit and comparing with the flat space definitions of mass and angular momentum (see e.g. Townsend's lectures). We emphasize that these integrals give the total mass and energy of the spacetime. This can come both from matter and from the gravitational field.

2.3 Killing horizons and surface gravity

Black holes and event horizons. A black hole is an asymptotically flat spacetime that contains a region which is not in the backward lightcone of future timelike infinity. The boundary of such region is called the event horizon. Put more simply, an event horizon is the boundary of a region in spacetime from behind which no causal signals can reach the observers sitting far away at infinity.



Null hypersurfaces. Consider a smooth function $f(x)$ of the spacetime coordinates x^μ . The level set $f(x) = \text{const}$ defines a *hypersurface*, that we denote by Σ . A vector $v = v^\mu \partial_\mu$ is *tangent* to Σ if it satisfies $v^\mu \partial_\mu f = 0$ (because f is constant along its level sets). The one-form

$$df = \partial_\mu f dx^\mu \quad (2.26)$$

is then *normal* to Σ , as it vanishes when acting on any tangent vector. Similarly, the vector field

$$\xi = g^{\mu\nu} \partial_\nu f \frac{\partial}{\partial x^\mu}, \quad (2.27)$$

is normal to Σ , as it is orthogonal to any tangent vector,

$$v \cdot \xi = v^\mu g_{\mu\nu} \xi^\nu = 0. \quad (2.28)$$

A *null hypersurface* \mathcal{N} is a hypersurface such that its normal vectors satisfy

$$\xi \cdot \xi = 0 \quad \text{on } \mathcal{N}. \quad (2.29)$$

In this case the normal vector ξ is also tangent to \mathcal{N} , as it satisfies $\xi^\mu \partial_\mu f = \xi^\mu \xi_\mu = 0$.

Killing horizons. A null hypersurface \mathcal{N} is said a *Killing horizon* if there exists a Killing vector field ξ that is normal to \mathcal{N} .

We are interested in Killing horizons because the event horizon of a stationary, asymptotically flat black hole is typically a Killing horizon.³ (The converse is not true, for instance in Minkowski space the Killing vector $\xi = x\partial_t + t\partial_x$ has norm $\xi_\mu \xi^\mu = -x^2 + t^2$ and thus is null at the surfaces $x = \pm t$, which however are not event horizons.) The Killing vector field associated with a Killing event horizon is a combination of the Killing vector $K = \partial_t$ generating time translations at infinity, and of the rotational Killing vector $\tilde{K} = \partial_\phi$, and can be written as

$$\xi = \partial_t + \Omega_H \partial_\phi, \quad (2.30)$$

where Ω_H is a constant called the *angular velocity* of the horizon. In the static case, $\xi = \partial_t$. Ω_H is interpreted as the angular velocity of the black hole in the sense that any test body dropped into it, as it approaches the horizon ends up circumnavigating it at this angular velocity, $\left. \frac{d\phi}{dt} \right|_{r \rightarrow r_+} = \Omega_H$.

Surface gravity. To every Killing horizon we can associate a quantity called *surface gravity*.

Since $\xi \cdot \xi = 0$ identically on \mathcal{N} , the gradient $\nabla_\mu(\xi \cdot \xi)$ is normal to \mathcal{N} , and therefore proportional to ξ at each point on \mathcal{N} . It follows that there exists a function κ , called the surface gravity of the Killing horizon, such that

$$\nabla_\mu(\xi \cdot \xi) = -2\kappa \xi_\mu \quad \text{on } \mathcal{N}. \quad (2.31)$$

Using the Killing equation (2.17), this can be rearranged as

$$\xi^\nu \nabla_\nu \xi^\mu = \kappa \xi^\mu \quad \text{on } \mathcal{N}. \quad (2.32)$$

³See e.g. Section 6.3 of Carroll's book for details.

This is the geodesic equation, where κ measures the failure of the integral curves of ξ to be affinely parameterized.⁴

A useful formula for the surface gravity in terms of a scalar equation is

$$\kappa^2 = -\frac{1}{2}\nabla^\mu\xi^\nu\nabla_\mu\xi_\nu \quad \text{on } \mathcal{N}. \quad (2.33)$$

This is derived as follows. Since ξ is normal to \mathcal{N} , by Frobenius theorem it satisfies

$$\xi_{[\mu}\nabla_\nu\xi_{\rho]} = 0. \quad (2.34)$$

Using the Killing equation $\nabla_{(\mu}\xi_{\rho)} = 0$, this equation can be rearranged as

$$\xi_\rho\nabla_\mu\xi_\nu = -2\xi_{[\mu}\nabla_{\nu]}\xi_\rho. \quad (2.35)$$

Multiplying by $\nabla^\mu\xi^\nu = \nabla^{[\mu}\xi^{\nu]}$ we obtain

$$\begin{aligned} \xi_\rho\nabla^\mu\xi^\nu\nabla_\mu\xi_\nu &= -2\xi_\mu\nabla^\mu\xi^\nu\nabla_\nu\xi_\rho \\ &= -2\kappa\xi^\nu\nabla_\nu\xi_\rho \\ &= -2\kappa^2\xi_\rho, \end{aligned} \quad (2.36)$$

thus proving (2.33).

Let us show that κ is constant on orbits of ξ . Take a vector v tangent to \mathcal{N} . Since (2.33) holds everywhere on \mathcal{N} , we can write on \mathcal{N}

$$v^\rho\nabla_\rho\kappa^2 = -\nabla^\mu\xi^\nu v^\rho\nabla_\rho\nabla_\mu\xi_\nu = -\nabla^\mu\xi^\nu v^\rho R_{\nu\mu\rho\sigma}\xi^\sigma, \quad (2.37)$$

where in the second equality we used property (2.18) of Killing vectors. Since ξ is also tangent, we can choose $v = \xi$, which gives

$$\xi^\rho\nabla_\rho\kappa^2 = -\nabla^\mu\xi^\nu R_{\nu\mu\rho\sigma}\xi^\rho\xi^\sigma = 0. \quad (2.38)$$

One can actually show that $v^\rho\nabla_\rho\kappa^2 = 0$ for every tangent vector, namely that κ is constant over the horizon. See e.g. Wald's book, Chapter 12.5 (pp. 333–334), for a proof.

Normalization of κ . Note that if \mathcal{N} is a Killing horizon for a Killing vector field ξ with surface gravity κ , then it is also a Killing horizon for $c\xi$ with surface gravity $c\kappa$, where c is any non-zero constant. This shows that the surface gravity is not an intrinsic property of the Killing horizon, it also depends on the normalization of ξ . While there is no natural normalization of ξ on \mathcal{N} (since it is null there), in a stationary, asymptotically flat spacetime we conventionally normalize the generator of time translations $K = \partial_t$ so that $K^\mu K_\mu = -1$ at spatial infinity; the sign is fixed by requiring that K is future-directed. This also fixes the normalization of $\xi = K + \Omega_H \tilde{K}$.

⁴An affine parameter λ is a parameter related to the proper time τ by an affine transformation, $\lambda = a\tau + b$.

Physical interpretation. As we will see, the main reason why we are interested in the surface gravity is that it provides the Hawking temperature of the black hole, which is a quantum effect. However, even in classical GR the surface gravity has a physical meaning.

In a static, asymptotically flat spacetime, the surface gravity is the acceleration of a particle at rest on the horizon, as measured by a static observer at infinity. The acceleration of a static observer near the horizon tends to infinity, but the redshift factor that relates this to the acceleration measured from infinity goes to zero. The surface gravity arises from the limiting value of the product of these two quantities, with the result typically being finite. When the spacetime is not static, this physical interpretation does not hold, but the surface gravity is still well-defined.

Let us check the above claims. Consider a static particle in a spacetime containing a static black hole. By definition, a static particle has four-velocity U^μ proportional to the time-translation Killing vector field, $K^\mu = V(x)U^\mu$, for some function $V(x)$. This function is called the “redshift factor”. Recalling that the four-velocity satisfies $U_\mu U^\mu = -1$, clearly we have $V = \sqrt{-K_\mu K^\mu}$. This ranges from 0 at the horizon to 1 at infinity.

Now consider the four-acceleration $a^\mu = U^\nu \nabla_\nu U^\mu$. We compute:

$$a_\mu = \frac{K^\nu}{V} \nabla_\nu \left(\frac{K_\mu}{V} \right) = \frac{K^\nu}{V^2} \nabla_\nu K_\mu - \frac{1}{V^3} K_\mu K^\nu \nabla_\nu V. \quad (2.39)$$

Using the Killing equation, the first term gives

$$\frac{1}{V^2} K^\nu \nabla_\nu K_\mu = -\frac{1}{V^2} K^\nu \nabla_\mu K_\nu = -\frac{1}{2V^2} \nabla_\mu (K_\nu K^\nu) = \frac{\nabla_\mu V}{V}, \quad (2.40)$$

while the second term vanishes, since

$$K^\nu \nabla_\nu V \propto K^\nu K^\rho \nabla_\nu K_\rho = K^\nu K^\rho \nabla_{(\nu} K_{\rho)} = 0. \quad (2.41)$$

We have thus found that the four-acceleration is given by

$$a_\mu = \frac{\nabla_\mu V}{V}, \quad (2.42)$$

and thus its magnitude is $a = \sqrt{a_\mu a^\mu} = V^{-1} \sqrt{\nabla_\mu V \nabla^\mu V}$. This is infinite at the horizon, as V vanishes there. But the acceleration as measured at infinity is redshifted by a factor of V , and reads

$$aV = \sqrt{\nabla_\mu V \nabla^\mu V}, \quad (2.43)$$

which is generically finite. One can check (see Wald, p. 332) that the square of this evaluated on the horizon agrees with our expression (2.33) for the surface gravity. Hence

$$\kappa = aV \quad \text{evaluated at the horizon.} \quad (2.44)$$

Let us apply the formulae above to Schwarzschild and evaluate its surface gravity. Recall that the metric is

$$ds^2 = -\left(1 - \frac{2GM}{r}\right) dt^2 + \left(1 - \frac{2GM}{r}\right)^{-1} dr^2 + r^2 d\Omega^2. \quad (2.45)$$

We have:

$$K = \partial_t \quad \Rightarrow \quad V = \sqrt{-K_\mu K^\mu} = \sqrt{-g_{tt}} = \sqrt{1 - \frac{2GM}{r}}, \quad (2.46)$$

so the acceleration is given by

$$a_\mu = \frac{\nabla_\mu V}{V} = \frac{GM}{r^2 \left(1 - \frac{2GM}{r}\right)} \delta_\mu^r, \quad (2.47)$$

and its magnitude is

$$a = \sqrt{a_\mu a^\mu} = \frac{GM}{r^2 \sqrt{1 - \frac{2GM}{r}}}. \quad (2.48)$$

It follows that

$$V a = \frac{GM}{r^2}. \quad (2.49)$$

The surface gravity is given by evaluating this quantity at the horizon $r = 2GM$, so

$$\kappa = \frac{1}{4GM}. \quad (2.50)$$

Notice that the surface gravity is inversely proportional to the mass, so it is large for small black holes, and vice-versa.

2.4 Generalized Smarr formula

Let us derive a relation between the mass, the horizon area, the angular momentum (and the electric charge) of a stationary, axisymmetric, asymptotically flat spacetime containing a black hole [1].

The Killing vector associated to the Killing horizon is $\xi = K + \Omega_H \tilde{K}$, where again K generates time translations and \tilde{K} is the angular Killing vector. The corresponding Komar conserved charge is a combination of the mass and the angular momentum of the spacetime:

$$Q_\xi = -\frac{1}{8\pi} \int_{S_\infty^2} *d\xi = -\frac{1}{8\pi} \int_{S_\infty^2} *dK - \frac{\Omega_H}{8\pi} \int_{S_\infty^2} *d\tilde{K} = M - 2\Omega_H J. \quad (2.51)$$

We can also evaluate Q_ξ in another way. Let Σ be a spacelike hypersurface intersecting the horizon, H , on a two-sphere S_H^2 , which together with the two-sphere S_∞^2 at spatial infinity

forms the boundary of Σ . Using the Stokes theorem we have:

$$\begin{aligned} Q_\xi &= -\frac{1}{8\pi} \int_{S_H^2} *d\xi - \frac{1}{8\pi} \int_\Sigma d * d\xi \\ &= -\frac{1}{8\pi} \int_{S_H^2} *d\xi + 2 \int_\Sigma (T_{\mu\nu} - \frac{1}{2}g_{\mu\nu}T) \xi^\nu * dx^\mu, \end{aligned} \quad (2.52)$$

where in the last step we used (2.20), (2.21). The integral over S_H^2 may be regarded as the contribution of the hole, while the one over Σ is a combination of the mass and angular momentum of the matter and radiation outside the horizon. In order to treat the integral over S_H^2 , we observe that the volume form on S_H^2 can be written as

$$\text{vol}_{S_H^2} = *(n \wedge \xi) \quad (2.53)$$

evaluated at the horizon. Here n^μ is another null vector normal to S_H^2 , normalized so that $n \cdot \xi = -1$. Hence

$$\begin{aligned} \int_{S_H^2} *d\xi &= \frac{1}{2} \int_{S_H^2} \text{vol}_{S_H^2} (*(n \wedge \xi))^{\mu\nu} (*d\xi)_{\mu\nu} \\ &= 2 \int_{S_H^2} \text{vol}_{S_H^2} n^\nu \xi^\mu \nabla_\mu \xi_\nu \\ &= -2\kappa A, \end{aligned} \quad (2.54)$$

where in the first step we project over the horizon and in the last step we used (2.32) together with the fact that κ is constant over the horizon, and $A = \int_{S_H^2} \text{vol}_{S_H^2}$ is the area of the horizon. Plugging this in (2.52) and comparing with (2.51), we arrive at

$$M = \frac{\kappa A}{4\pi} + 2\Omega_H J + 2 \int_\Sigma (T_{\mu\nu} - \frac{1}{2}g_{\mu\nu}T) \xi^\nu * dx^\mu. \quad (2.55)$$

If we are in pure GR, $T_{\mu\nu} = 0$. Then our spacetime is the Kerr black hole and the formula reads

$$M = \frac{\kappa A}{4\pi} + 2\Omega_H J. \quad (2.56)$$

This is Smarr's formula for the mass of a Kerr black hole.

Exercise. If we consider the Einstein-Maxwell theory (see (2.59) below), the energy-momentum tensor is the one of the electromagnetic field, $F_{\mu\nu}$. Show that in this case the formula becomes

$$M = \frac{\kappa A}{4\pi} + 2\Omega_H J + \Phi_H Q, \quad (2.57)$$

where Φ_H is the co-rotating electric potential on the horizon, which for a gauge field vanishing at infinity is defined as

$$\Phi_H = -\xi^\mu A_\mu \quad \text{evaluated at the horizon.} \quad (2.58)$$

This equals the line integral of the hole's electric field from infinity to the horizon (and is independent of the position at the horizon).

2.5 The Kerr-Newman solution

Let us see how the concepts discussed above work in a concrete example. Consider the Einstein-Maxwell theory in four dimensions,

$$S = \frac{1}{16\pi} \int d^4x \sqrt{-g} (R - F_{\mu\nu} F^{\mu\nu}), \quad (2.59)$$

where $F = dA$, A being an Abelian gauge field. The Einstein and Maxwell equations are

$$\begin{aligned} R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R &= 2F_{\mu\rho} F_{\nu}{}^\rho - \frac{1}{2} g_{\mu\nu} F_{\rho\sigma} F^{\rho\sigma}, \\ \nabla^\mu F_{\mu\nu} &= 0. \end{aligned} \quad (2.60)$$

The most general stationary black hole solution to this theory⁵ is given by the Kerr-Newman solution. The metric and gauge field in Boyer-Lindquist coordinates read

$$\begin{aligned} ds^2 &= -\frac{\Delta - a^2 \sin^2 \theta}{\Sigma} dt^2 - 2a \frac{r^2 + a^2 - \Delta}{\Sigma} \sin^2 \theta dt d\phi \\ &+ \frac{(r^2 + a^2)^2 - \Delta a^2 \sin^2 \theta}{\Sigma} \sin^2 \theta d\phi^2 + \frac{\Sigma}{\Delta} dr^2 + \Sigma d\theta^2, \end{aligned} \quad (2.61)$$

$$A = -\frac{1}{\Sigma} [Qr(dt - a \sin^2 \theta d\phi) + P \cos \theta (a dt - (r^2 + a^2) d\phi)], \quad (2.62)$$

where

$$\Sigma = r^2 + a^2 \cos^2 \theta, \quad \Delta = r^2 - 2Mr + a^2 + P^2 + Q^2, \quad (2.63)$$

and M, a, P, Q are parameters. It will be convenient to write the quadratic polynomial $\Delta(r)$ in terms of its roots,

$$\Delta = (r - r_+)(r - r_-), \quad (2.64)$$

where

$$r_\pm = M \pm \sqrt{M^2 - (a^2 + P^2 + Q^2)}, \quad (2.65)$$

⁵The statement that this is the most general stationary black hole solution extends to other theories with matter couplings, for some details see Wald's book, Section 12.3.

that it may sometimes be convenient to express as

$$r_+ + r_- = 2M, \quad r_+ r_- = a^2 + P^2 + Q^2. \quad (2.66)$$

We can make some remarks:

- For $a = 0$, the solution reduces to the Reissner-Nordström solution. For $P = Q = 0$, the gauge field vanishes and the metric reduces to Kerr. For $a = P = Q = 0$, we obtain Schwarzschild. In these lectures we will often take one of these limits, depending on convenience.

- At first order near to $r \rightarrow \infty$, the metric reads

$$ds^2 \sim -dt^2 + dr^2 + r^2 (d\theta^2 + \sin^2 \theta d\phi^2). \quad (2.67)$$

This is the metric of Minkowski space in polar coordinates (r, t, θ, ϕ) , with θ, ϕ parameterizing an S^2 provided we take $0 < \theta < \pi$, $\phi \sim \phi + 2\pi$. Hence the space is asymptotically flat. Asymptotically it is also stationary and axisymmetric. Indeed, it is clear that $K = \partial/\partial t$ and $\tilde{K} = \partial/\partial \phi$ are (commuting) Killing vector, as the metric components don't depend on the t, ϕ coordinates. Moreover, for sufficiently large r , $K = \partial/\partial t$ is timelike, while $\tilde{K} = \partial/\partial \phi$ is spacelike. This is enough for obtaining conserved charges via the Komar integrals.

Also note that the discrete transformation $t \rightarrow -t$ is not a symmetry, implying that the solution is not static. Instead, $t \rightarrow -t$ together with $\phi \rightarrow -\phi$, which reverses both the direction of time and the sense of rotation, *is* a symmetry; this is a feature of stationary solutions.

Exercise. Using a computer algebra program, check that the metric (2.61) and the gauge field (2.62) solve the Einstein and Maxwell equations. Check that applying the definitions of electric charge, magnetic charge, Komar mass and Komar angular momentum given above, one obtains $Q, P, M_{\text{Komar}} = M, J_{\text{Komar}} = aM$, respectively.

- Start from infinity and move towards lower values of r . For very large r , the polynomial $\Delta(r)$ is positive, and its value decreases while we reduce r . At some point we will reach $\Delta = 0$, where something special happens as g_{rr} blows up. In order to understand this better, let us look at the metric on the two-dimensional hypersurfaces at constant r and constant θ . This is

$$g_{2d} = g_{tt} dt^2 + 2g_{t\phi} dt d\phi + g_{\phi\phi} d\phi^2, \quad (2.68)$$

where g_{tt} , $g_{t\phi}$ and $g_{\phi\phi}$ can be read off from (2.61). This is non-degenerate and Lorentzian as long as

$$\det(g_{2d}) = g_{tt}g_{\phi\phi} - g_{t\phi}^2 < 0. \quad (2.69)$$

The condition

$$\det(g_{2d}) = g_{tt}g_{\phi\phi} - g_{t\phi}^2 = 0 \quad (2.70)$$

defines a null hypersurface. For the Kerr-Newman solution, we have

$$\det(g_{2d}) = -\Delta \Sigma \sin^2 \theta, \quad (2.71)$$

so as long as $\Delta > 0$ the metric is indeed Lorentzian, while at $\Delta = 0$ it is degenerate. This locus is in fact a Killing horizon. The Killing vector becoming null at $r = r_+$ is

$$\xi = \frac{\partial}{\partial t} + \Omega_H \frac{\partial}{\partial \phi}, \quad (2.72)$$

where we defined

$$\Omega_H = - \left. \frac{g_{t\phi}}{g_{\phi\phi}} \right|_{r_+} = \frac{a}{r_+^2 + a^2}. \quad (2.73)$$

This is easily checked by computing the squared norm $\xi \cdot \xi$ using (2.68), and recalling that at $r = r_+$ the condition (2.70) is satisfied. The constant Ω_H is the angular velocity of the event horizon (with respect to a non-rotating frame at infinity).⁶

Using $a = J/M$ and the expression (2.65) for r_+ , one can find an expression for Ω_H in terms of the black hole charges. In particular, for the Kerr black hole ($Q = P = 0$) we obtain

$$\Omega_H = \frac{J}{2M (M^2 + \sqrt{M^4 - J^2})}. \quad (2.74)$$

The roots r_{\pm} of Δ denote the positions of the outer and inner Killing horizons, the former being the event horizon. Introducing null coordinates, one can see that these are just coordinate singularities, and the metric and gauge field are actually smooth there.

- The other special locus is $\Sigma = 0$. By computing $R_{\mu\nu\rho\sigma}R^{\mu\nu\rho\sigma}$, one finds that, contrary to $\Delta = 0$, this is a curvature singularity as long as $M \neq 0$. Since $\Sigma = r^2 + a^2 \cos^2 \theta$, the condition $\Sigma = 0$ is equivalent to $r = 0$ and $\theta = \pi/2$. This locus has the shape of a ring, so we have a curvature ring singularity (a “ringularity”!). In order to see this, first restrict the metric (2.61) (with $P = Q = 0$) to $r = 0$. The spatial part parameterized by θ, ϕ describes a disk, with $\theta = \pi/2$ being its boundary, that is a ring.

One can show that it is possible to pass through the singularity and reach a region where $r < 0$. This region contains timelike curves near the singularity, because $g_{\phi\phi}$ becomes negative for small negative r , and the orbits of the vector ∂_{ϕ} are closed. Closed timelike curves are considered a causal pathology.

⁶One could repeat the same steps for the inner horizon at $r = r_-$.

- We will always assume

$$M^2 \geq a^2 + P^2 + Q^2, \quad M > 0, \quad (2.75)$$

so that the roots (2.65) are real and positive. If this condition is not met, the curvature singularity at $r = 0$ is not screened by a horizon and we would have a *naked singularity*. Naked singularities are believed to be non-physical.

- Let us check the Smarr's relation for the Kerr black hole ($Q = P = 0$), given by (2.56). Evaluating (2.33), we find for the surface gravity

$$\kappa = \frac{r_+ - r_-}{2(r_+^2 + a^2)}. \quad (2.76)$$

Note that this does not depend on the horizon coordinates, in agreement with the general result stating that the surface gravity is constant over the horizon. The area of the event horizon is

$$A = \int_{S_H^2} \text{vol}_H = \int_0^\pi d\theta \int_0^{2\pi} d\phi \sqrt{g_{\theta\theta}g_{\phi\phi}}|_{r_+} = 4\pi (r_+^2 + a^2), \quad (2.77)$$

where $g_{\theta\theta}$, $g_{\phi\phi}$ are read from (2.61), recalling that $\Delta(r_+) = 0$. Then we have

$$\frac{\kappa A}{4\pi} = \frac{r_+ - r_-}{2} = r_+ - M, \quad 2\Omega_H J = \frac{2a^2 M}{r_+^2 + a^2} = \frac{a^2}{r_+}, \quad (2.78)$$

where we recalled that $J = aM$ in the very last step we used $\Delta = 0 \Leftrightarrow r_+^2 + a^2 = 2Mr_+$. Hence

$$\frac{\kappa A}{4\pi} + 2\Omega_H J = \frac{r_+^2 - Mr_+ + a^2}{r_+} = M, \quad (2.79)$$

that is the Smarr relation.

Exercise. Check that for $Q \neq 0$, $P = 0$, the electric potential, defined as $A = -\xi \cdot A|_{r_+}$ is

$$\Phi_H = \frac{Q r_+}{r_+^2 + a^2}. \quad (2.80)$$

Also check that the generalized Smarr relation holds in the form (2.57). Note that this is constant over the horizon.

2.6 Ergosphere

Consider the Kerr solution ($Q = P = 0$). A peculiar feature is that the Killing vector $K = \partial_t$, that asymptotically is timelike, becomes null and then spacelike before we reach $r = r_+$. Indeed

$$K_\mu K^\mu = g_{tt} = \frac{a^2 \sin^2 \theta - \Delta}{\Sigma} \quad (2.81)$$

is < 0 for $r \rightarrow \infty$, but is ≥ 0 at $r = r_+$, where $\Delta = 0$ (it vanishes only at $\theta = 0, \pi$). The vector becomes null on the hypersurface

$$r^2 - 2Mr + a^2 \cos^2 \theta = 0. \quad (2.82)$$

The region included between this hypersurface and the outer horizon $r = r_+$ is called the *ergosphere*.

In the ergosphere, all observers are forced to rotate in the direction of rotation of the black hole. To see this, let us set ourselves on the equatorial plane $\theta = \pi/2$ and emit a photon in the ϕ direction. At the precise instant when it is emitted, the photon has no components along the r or θ directions, so the condition for a null trajectory is

$$g_{tt} dt^2 + 2g_{t\phi} dt d\phi + g_{\phi\phi} d\phi^2 = 0 \quad (2.83)$$

implying that

$$\frac{d\phi}{dt} = -\frac{g_{t\phi}}{g_{\phi\phi}} \pm \sqrt{\left(\frac{g_{t\phi}}{g_{\phi\phi}}\right)^2 - \frac{g_{tt}}{g_{\phi\phi}}}. \quad (2.84)$$

Evaluating this at the outer boundary of the ergosphere, where $g_{tt} = 0$, we obtain the two solutions

$$\frac{d\phi}{dt} = 0, \quad \frac{d\phi}{dt} = -2 \frac{g_{t\phi}}{g_{\phi\phi}}. \quad (2.85)$$

The second solution has the same sign as a , so it describes a photon directed in the same direction as the black hole's rotation. The vanishing solution describes a photon emitted in the opposite direction: it "tries" to travel against the hole's rotation, but the best it can achieve is that its instantaneous velocity is zero. This phenomenon happens within the ergosphere and is known as *dragging of inertial frames*. Any massive particle will not be able to do better than the photon, and will be dragged in the sense of rotation of the black hole.

If we go down to $r = r_+$ and again emit a photon, we find

$$\left. \frac{d\phi}{dt} = -\frac{g_{t\phi}}{g_{\phi\phi}} \right|_{r_+} = \Omega_H \quad (2.86)$$

because the argument of the square root is proportional to Δ and thus vanishes in $r = r_+$. This means that the photon must move in the ϕ direction at the same velocity as the black hole.

2.7 Penrose process for energy extraction

The Penrose process is a mechanism allowing to extract energy from a rotating black hole.

Consider a spaceship moving along a geodesic in the ergoregion of a Kerr black hole. The four-momentum is

$$p^\mu = mu^\mu, \quad \text{where } u^\mu = \frac{dx^\mu}{d\tau} \text{ is the four-velocity.} \quad (2.87)$$

The energy and angular momentum of the particle are the conserved quantities associated with the Killing vectors K and \tilde{K} ,⁷

$$\begin{aligned} E &= -K_\mu p^\mu, \\ L &= \tilde{K}_\mu p^\mu. \end{aligned} \quad (2.88)$$

The definition of the energy contains a minus sign because at infinity both K^μ and p^μ are timelike, hence their product is negative, while we want the energy to be positive.

Imagine the spaceship launches a heavy bullet. By the conservation of four-momentum, we have

$$p^\mu = p_1^\mu + p_2^\mu, \quad \Rightarrow \quad E = E_1 + E_2, \quad (2.89)$$

where “1”=ship, “2” = bullet. Since we are in the ergoregion, the vector $K = \partial_t$ is spacelike and it is possible to arrange the launch of the bullet so that its energy is negative, $E_2 < 0$. Then $E_1 = E + |E_2| > E$. It can be shown that the body with $E_2 < 0$ eventually falls into the black hole, while the spaceship can safely exit the ergoregion. So by traveling in and out the ergoregion the spaceship has acquired energy! Energy is conserved in the process as the black hole absorbs a negative energy and its total mass decreases.

However there is a limit to the energy that can be extracted in this way. Consider the Killing vector becoming null at the horizon of Kerr, $\xi = \partial_t + \Omega_H \partial_\phi$. The statement that the body with momentum p_2^μ crosses the horizon “traveling forward in time” is

$$\xi_\mu p_2^\mu \leq 0 \quad \Leftrightarrow \quad \Omega_H L_2 \leq E_2. \quad (2.90)$$

⁷Recall that the projection of any Killing vector k^μ along a geodesic is conserved, $k_\mu u^\mu = \text{constant}$ along the geodesic, that is $u^\nu \nabla_\nu (k_\mu u^\mu) = 0$. This immediately follows from the Killing equation $\nabla_{(\nu} k_{\mu)} = 0$ and the geodesic equation $u^\nu \nabla_\nu u^\mu = 0$.

After absorbing the energy E_2 and the angular momentum L_2 of the bullet, the black hole will settle to a Kerr black hole with mass $M + \delta M$, with $\delta M = E_2 < 0$, and angular momentum $J + \delta J$, with $\delta J = L_2$, such that

$$\Omega_H \delta J \leq \delta M. \quad (2.91)$$

This inequality is telling us that the loss in angular momentum is somehow more important than the mass loss. In order to quantify this more precisely, following Christodoulou [2] we introduce the *irreducible mass*

$$M_{\text{irr}} = \sqrt{\frac{1}{2} \left(M^2 + \sqrt{M^4 - J^2} \right)}, \quad (2.92)$$

satisfying $M_{\text{irr}} \leq M$, with the equality holding only when $J = 0$. It takes only a few steps to show that

$$\begin{aligned} \delta M_{\text{irr}}^2 &= \frac{1}{2} \left(2M\delta M + \frac{2M^3\delta M - J\delta J}{\sqrt{M^4 - J^2}} \right) \\ &= \frac{J}{2\Omega_H\sqrt{M^4 - J^2}} (\delta M - \Omega_H \delta J) \geq 0, \end{aligned} \quad (2.93)$$

where in the second line we have used Eq. (2.74) for the angular velocity. So we have found that in a Penrose process the irreducible mass cannot decrease,

$$\delta M_{\text{irr}} \geq 0, \quad (2.94)$$

which explains the name given to it. The process starts with $M_{\text{irr}} < M$, and necessarily leads to an increase of M_{irr} . The maximum value that can be reached is $M_{\text{irr}} = M$, when the black hole angular momentum J goes to zero and one cannot extract further energy through the Penrose process.

Note that the irreducible mass is related to the area of the event horizon (2.77) as

$$A = 16\pi M_{\text{irr}}^2 \quad (2.95)$$

(check it as an exercise). Hence we have found that in a Penrose process necessarily

$$\delta A \geq 0. \quad (2.96)$$

This is a manifestation of a much more general theorem, due to Hawking (see the next section).

Using (2.95) into (2.93) and noting that $\kappa = \sqrt{M^4 - J^2} \Omega_H / J$, it also follows that

$$\delta A = \frac{8\pi}{\kappa} (\delta M - \Omega_H \delta J), \quad (2.97)$$

that we can rearrange as

$$\delta M = \frac{\kappa}{8\pi} \delta A + \Omega_H \delta J. \quad (2.98)$$

Eqs. (2.96), (2.98) remind us of the second and first law of thermodynamics, respectively, provided we identify A as being proportional to some sort of black hole entropy, and κ as being proportional to a temperature. Historically, this gave a first hint that black holes obey the laws of thermodynamics. In the next section we give more substance to such analogy.

3 Black hole thermodynamics

In this section we discuss how using just GR one can establish four laws of black hole mechanics, that are formally analogous to the laws of thermodynamics. We will also start seeing that in fact this is not just a formal analogy, in particular we will see that black holes do have a physical entropy and really behave as thermodynamical systems.

3.1 Why should black holes carry an entropy?

Bekenstein was the first to propose that black holes should carry an entropy, and that this should be proportional to the area of the event horizon [3, 4]. Two arguments supporting this intuition are the following:

- Black holes are formed from the collapse of matter, which carries entropy. However, the matter that has contributed to form a black hole is not visible from an observer watching from outside the event horizon. So this observer must conclude either that the entropy disappears in the formation and growth of black holes, and thus that the second principle of thermodynamics is violated, or that the black holes themselves carry entropy. This issue can be summarized with the question [attributed to Wheeler, Bekenstein’s advisor]: “what happens if we throw a cup of tea into a black hole?”.
- A bit more quantitatively, let us imagine to throw “quanta” into a Schwarzschild black hole. The number of states goes as e^N , so the entropy is proportional to N . In order to fit, the size of the quanta should be at most the Schwarzschild radius r_s , so their energy should be at least $1/r_s$. For a black hole of mass $M \sim r_s/G$, the change in entropy is at most

$$dS \sim dN \sim r_s dM \sim \frac{r_s dr_s}{G} \sim \frac{dA}{G}. \quad (3.1)$$

This heuristic argument is a first hint that the black hole entropy may be proportional to the area.

- In general relativity, black hole solutions are fully characterized by few conserved quantities, such as the mass, the angular momentum, and the electric charge. This is Wheeler’s famous statement that “black holes have no hair”. However there are many ways of forming a black hole with assigned values of these quantities. From this perspective, black holes are macroscopic thermodynamic objects with many microstates, corresponding to the different possible ways of forming the same macroscopic solution. Enumerating these microstates leads to the entropy.

3.2 The laws of black hole mechanics

We now present the four laws of black hole mechanics [1] and discuss their analogy with thermodynamics. Let us start from the most suggestive one:

Second law. *In any physical process, the area A of the event horizon does not decrease,*

$$\Delta A \geq 0. \tag{3.2}$$

This is Hawking’s celebrated area theorem [5]. This theorem assumes validity of cosmic censorship, i.e. that singularities which occur in gravitational collapse are always cloaked behind an event horizon. It also assumes that the energy-momentum tensor of the matter fields obeys the weak energy condition. We will not prove it here; see e.g. Wald’s book.

Motivated by the idea that black holes should carry an entropy, Bekenstein pointed out the analogy of Hawking’s black hole area theorem with the second law of thermodynamics, which states that in physical processes the entropy does not decrease:

$$\Delta S \geq 0. \tag{3.3}$$

This leads to argue that the black hole entropy S is a monotonic function of A/ℓ_P^2 , where the Planck length ℓ_P is introduced for dimensional reasons. In a moment we will see that the simplest assumption that the black hole entropy is just proportional to A/ℓ_P^2 is the correct one. This is a surprising and far-reaching observation. It is surprising because the entropy usually is an extensive quantity growing with the volume, not with the area. It is far-reaching for many reasons, one being that it was crucial to develop the holographic principle, that plays a central role in our contemporary understanding of quantum gravity.

Generalized second law. If one considers the ordinary entropy in a region outside a black hole, this may well decrease as long as matter falls into the black hole. This led Bekenstein

to formulate [3, 4] a generalization of the second law of thermodynamics, stating that *the sum of ordinary entropy outside black holes and the total black hole entropy never decreases*. According to this principle, the increase in black hole entropy must more than compensate for the disappearance of ordinary entropy from the outside region. This principle has been verified in a number of examples.

Zeroth law. *The surface gravity is constant over a Killing horizon.* In ordinary thermodynamics, the temperature is a quantity that is everywhere the same in a system at equilibrium, and this fact is expressed by the zeroth law of thermodynamics. So in the analogy between black hole mechanics and thermodynamics, we could think of the surface gravity as a temperature.

First law. We would like to test the idea that black holes have an entropy proportional to the horizon area. Both in black hole physics and in ordinary thermodynamics energy is conserved. In ordinary thermodynamics, conservation of the energy is expressed by the first principle, which says that in an infinitesimal transformation

$$dE = TdS + dW, \quad (3.4)$$

where dW is the work done on the system; for instance $dW = p dV$. When the system rotates with angular velocity Ω and is charged up to an electric potential Φ , the changes dJ and dQ in its angular momentum and electric charge contribute to the work done on it in such a way that

$$dE = TdS + \Omega dJ + \Phi dQ. \quad (3.5)$$

In black hole mechanics one has an analog statement, which generalizes what we have seen in Section 2.7 while discussing the Penrose process:

If a stationary black hole of mass M , angular momentum J and electric charge Q is perturbed so that it settles down to another black hole of mass $M + \delta M$, angular momentum $J + \delta J$ and charge $Q + \delta Q$, then

$$dM = \frac{\kappa}{8\pi} dA + \Omega_H dJ + \Phi_H dQ. \quad (3.6)$$

By comparison with (3.5), we are led to identify $T = \alpha \frac{\kappa}{8\pi}$ and $S = A/\alpha$, where α is some constant. So the hypothesis that the black hole entropy is just proportional to A and that the surface gravity provides a temperature seems good indeed.

Proof. Let us sketch a proof of (3.6) for $Q = 0$.⁸ Uniqueness theorems for the Kerr black hole imply that $M = M(A, J)$. In units such that $c = G = 1$, both A and J have dimensions of M^2 , so the function $M(A, J)$ must be homogeneous of degree $1/2$. Euler theorem of homogeneous functions then implies that

$$\begin{aligned} A \frac{\partial M}{\partial A} + J \frac{\partial M}{\partial J} &= \frac{1}{2} M \\ &= \frac{\kappa}{8\pi} A + \Omega_H J, \end{aligned} \tag{3.7}$$

where in the second line we used Smarr's formula. Rearranging, we have

$$A \left(\frac{\partial M}{\partial A} - \frac{\kappa}{8\pi} \right) + J \left(\frac{\partial M}{\partial J} - \Omega_H \right) = 0. \tag{3.8}$$

But A and J are free parameters, so

$$\frac{\partial M}{\partial A} = \frac{\kappa}{8\pi}, \quad \frac{\partial M}{\partial J} = \Omega_H, \tag{3.9}$$

which proves the statement.

Exercise. Consider our example of the Kerr-Newman black hole and check explicitly that (3.6) holds. This exercise was first done in [4].

Third law. There also exists a black hole analog of the third law of thermodynamics, although it is on less firm grounds. A formulation of the third law of thermodynamics states that a thermal system cannot reach zero temperature in a finite number of physical processes. A zero-temperature black hole, namely a black hole whose surface gravity vanishes, is an allowed solution to the equations of motion and is called *extremal*. In the case of Kerr-Newman, this condition corresponds to $M^2 = a^2 + Q^2 + P^2$. So the black hole counterpart of the third principle would be that *no physical process exists that allows to reach an exactly extremal black hole*. For the Kerr and electrically charged Kerr black holes, calculations have been done showing that the closer one gets to an extreme black hole, the harder it becomes to get a further step closer.

An alternate formulation of the third law of thermodynamics says that the entropy of a system approaches a constant value as the temperature approaches zero. By constant we mean that it should not depend on the intensive variables, such as the pressure, the electric potential or the magnetic field. In the statistical mechanics interpretation of the entropy, this

⁸This proof is due to Gibbons and is taken from Townsend, p. 113.

value is related to the number of microscopic ground states of the system. In the particular case where there is just one ground state the entropy vanishes at zero temperature. Extremal black holes in general have non-vanishing area and thus still carry macroscopic entropy; so in some sense we should think they are made of a very large number of ground states (more later).

Appearance of \hbar . Restoring all physical units, the formulae for the black hole entropy and temperature read

$$S = k_B \frac{A}{\alpha \ell_P^2}, \quad T = \frac{\hbar}{ck_B} \frac{\alpha \kappa}{8\pi}, \quad (3.10)$$

with $\ell_P^2 = G\hbar/c^3$ and α is just a numerical constant. The $1/\ell_P^2$ factor is motivated by dimensional analysis and the fact that there are no other scales in the problem. However this is not innocent at all: it brings in a factor of \hbar both in the entropy and in the temperature.

Note that the combination appearing in the first law (3.6) is $TdS = \frac{c^2}{8\pi G} \kappa dA$ and does not contain neither \hbar , nor k_B ; this is totally expected, since this relation has been derived in classical GR, which does not contain neither k_B , nor \hbar . The physical units of κ are those of an acceleration (not a temperature) while A is an area. The factor of $\frac{k_B}{\ell_P^2}$, that was introduced by Bekenstein's intuition, converts A and $\frac{c^2}{G}\kappa$ into thermodynamical quantities.

Some numbers. Note that an entropy proportional to ℓ_P^{-2} is huge compared to the entropy of ordinary matter systems. For a solar mass black hole, the Schwarzschild radius is $r_h = \frac{2GM}{c^2} \simeq 3 \cdot 10^3$ m, the area is $A = 4\pi r_h^2 \simeq 10^8$ m². The Planck length is $\ell_P \simeq 1.6 \cdot 10^{-35}$ m, so one obtains $\frac{A}{\ell_P^2} \sim 4 \cdot 10^{77}$, that is about twenty orders of magnitude larger than the thermodynamic entropy of the Sun. This shows that the entropy of a black hole is not just the thermodynamical entropy of the bodies that formed it. For a black hole of 10^6 solar masses, the entropy is $\sim 10^{90}$, that exceeds the thermodynamic entropy in the whole universe. This also means that the universe is in a low-entropy state, as the entropy could be made much larger by throwing more matter into black holes.

As already noticed, the appearance of \hbar rather indicates that the microstates responsible for the black hole entropy are quantum.

The numerical coefficient α in (3.10) remains undetermined at this stage. As we are going to see next, it is fixed to $\alpha = 4$ by Hawking's calculation showing that quantum particle creation effects result in a thermal emission of particles from a black hole at a temperature $T = \frac{\hbar\kappa}{2\pi}$.

4 QFT in curved spacetime and Hawking radiation

In classical GR, the analogy of black hole mechanics with thermodynamics is just formal. Indeed in GR black holes did not emit any radiation, so they should be regarded as bodies at absolute zero temperature. Moreover, the laws of black hole mechanics are mathematically exact consequences of GR, while the laws of thermodynamics are not fundamental, they only emerge once one considers systems with a very large number of degrees of freedom; so the analogy discussed in the previous section may seem accidental. However, the fact that the black hole temperature is proportional to \hbar suggests that the reason why it is not computable in GR is that it is entirely due to quantum effects.

An argument supporting the idea that black holes may radiate comes from the generalized second law. Indeed if black holes did not emit any radiation, then it would be easy to violate the generalized second law by simply considering a black hole immersed in a thermal bath at temperature lower than the formal temperature assigned to the hole. Indeed one would have $T_{\text{BH}} dS_{\text{BH}} + T_{\text{bath}} dS_{\text{bath}} = 0$, with $dS_{\text{BH}} > 0$ and $dS_{\text{bath}} < 0$. If $T_{\text{BH}} > T_{\text{bath}}$, then $T_{\text{BH}} (dS_{\text{BH}} + dS_{\text{bath}}) < 0$.

The breakthrough happened in 1974, when Hawking calculated particle creation effects for a body that collapses to a black hole, and discovered that a distant observer sees a thermal distribution of particles emitted at the temperature [6]

$$T = \frac{\hbar\kappa}{2\pi}. \quad (4.1)$$

So the black hole temperature is truly physical, and black hole thermodynamics is fully meaningful. In particular, if one placed a black hole in a radiation bath of temperature $T_{\text{bath}} < T_{\text{BH}}$, the black hole radiation would dominate over absorption, and there would be no violation of the generalized second law. The entropy $S_{\text{BH}} = A/4$ could now be interpreted as the physical entropy of the black hole, with the unknown constant in Bekenstein's original proposal now fixed by Hawking's computation of the temperature.

4.1 QFT in curved spacetime

Hawking radiation arises from studying QFT in curved spacetime. We do not need to quantize gravity to see it, we just need to consider quantum fields in the background of a black hole geometry, which is treated classically.

This is to some extent analog to the Schwinger effect in QED. The Schwinger effect consists of the production of an electron-positron pair out of the vacuum in the background of

a strong electric field, which is treated classically. Electron-positron pairs are spontaneously created in the vacuum, and the strong electric field separates them before they can annihilate with each other.

Quantum field theory in flat spacetime is based on Lorentz invariance. For instance, the Klein-Gordon equation for a real scalar field,

$$\eta^{\mu\nu} \partial_\mu \partial_\nu \varphi = m^2 \varphi, \quad (4.2)$$

admits plane wave solutions $e^{ik \cdot x}$, with $k^\mu k_\mu = -m^2$. Separating the positive and negative frequency waves, the general solution can be written as

$$\varphi(x) = \int \frac{d^3k}{(2\pi)^3 2\omega_k} \left(a_{\vec{k}} e^{ik \cdot x} + a_{\vec{k}}^\dagger e^{-ik \cdot x} \right)_{k^0 = \omega_k} \quad (4.3)$$

with $\omega_k = \sqrt{m^2 + |\vec{k}|^2}$. This respects Lorentz symmetry: different inertial observers may use different spacetime coordinates and thus perform different mode expansions for the same field, however these are simply related by a Lorentz transformation. One consequence is that all inertial observers will see the same vacuum state. The vacuum is defined as the Poincaré invariant state that is annihilated by half of the oscillators,

$$a_{\vec{k}} |0\rangle = 0. \quad (4.4)$$

With this definition, all inertial observers will agree on the number of particles contained in the vacuum.

In curved space things work differently. Making the minimal substitution $\eta^{\mu\nu} \rightarrow g^{\mu\nu}$, $\partial_\mu \rightarrow \nabla_\mu$, the Klein-Gordon equation becomes

$$\nabla^2 \varphi \equiv g^{\mu\nu} \nabla_\mu \partial_\nu \varphi = 0. \quad (4.5)$$

In general it is hard to find solutions to this equation. Moreover, it is not obvious how to separate modes of positive and negative frequency. In order to do this we need an isometry. Assume we have a Killing vector $K = K^\mu \partial_\mu$. Then one can show that this commutes with the Laplacian when acting on functions (you may verify this as an exercise),

$$[K, \nabla^2] f = 0. \quad (4.6)$$

We can introduce the inner product

$$(f_1, f_2) = i \int_\Sigma d^3x \sqrt{\gamma} n^\mu (\bar{f}_1 \partial_\mu f_2 - \partial_\mu \bar{f}_1 f_2), \quad (4.7)$$

where Σ is a spacelike hypersurface, n^μ is its unit normal vector and γ is the determinant of the induced metric. Since ∇^2 and iK are both self-adjoint, namely they are operators O satisfying $(Of_1, f_2) = (f_1, Of_2)$, they admit a complete set of common eigenfunctions,

$$\nabla^2 f = m^2 f, \quad iK^\mu \partial_\mu f = \nu f. \quad (4.8)$$

If K is timelike, we are entitled to call frequency its eigenvalue. Indeed in the Minkowski case we have $iK^\mu \partial_\mu = i\partial_t$, and on the plane wave $f = e^{ik \cdot x}$ it gives $iKf = \nu f$, with $\nu = k^0$. Notice that if f is an eigenfunction of positive frequency ν , then f^* is an eigenfunction of negative frequency $-\nu$.

Therefore in a spacetime admitting a timelike Killing vector we can expand our field in positive and negative frequency eigenfunctions of the Laplacian as

$$\varphi = \int_0^\infty d\nu (a_\nu f_\nu + a_\nu^\dagger f_\nu^*), \quad (4.9)$$

with $[a_\nu, a_{\nu'}^\dagger] = 2\pi\delta(\nu - \nu')$. Here the eigenfunctions are orthonormal with respect to the inner product (4.7), namely they satisfy $(f_\nu, f_{\nu'}) = 2\pi\delta(\nu - \nu')$.

In this situation, however, two different observers may choose two different timelike Killing vectors to define their frequencies, and these are in general not equivalent. So the two observers will have two truly different positive and negative frequency mode expansions. Namely,

$$\varphi = \underbrace{\int_0^\infty d\nu (a_\nu f_\nu + a_\nu^\dagger f_\nu^*)}_{\text{observer 1}} = \underbrace{\int_0^\infty d\omega (b_\omega g_\omega + b_\omega^\dagger g_\omega^*)}_{\text{observer 2}}, \quad (4.10)$$

with $[b_\omega, b_{\omega'}^\dagger] = 2\pi\delta(\omega - \omega')$. Since the eigenfunctions g_ω and f_ν both form a complete set, they can be expanded one into the other, for instance

$$f_\nu = \int_0^\infty \frac{d\omega}{2\pi} (\alpha_{\omega\nu} g_\omega + \beta_{\omega\nu}^* g_\omega^*). \quad (4.11)$$

This leads to an expansion of one set of raising and lowering operators into the other, for instance,

$$b_\omega = \int_0^\infty \frac{d\nu}{2\pi} (\alpha_{\omega\nu} a_\nu + \beta_{\omega\nu} a_\nu^\dagger). \quad (4.12)$$

The coefficients α and β are called *Bogoliubov coefficients*.

Since the two observers use different Killing vectors to describe time translations, they will define different Hamiltonians, and therefore they will in general identify different states as the minimum energy state, that is the vacuum. In particular, the state that is identified as the empty vacuum state for the first observer, may be full of particles for the second

observer. Indeed if the state $|\psi\rangle$ satisfies $a_\nu|\psi\rangle = 0$ for all $\nu > 0$ and is thus identified as the vacuum by the first observer, then the particle occupation number for the second observer will be

$$\begin{aligned} \langle\psi|b_\omega^\dagger b_{\omega'}|\psi\rangle &= \int \frac{d\nu}{2\pi} \int \frac{d\nu'}{2\pi} \beta_{\omega\nu}^* \beta_{\omega'\nu'} \underbrace{\langle\psi|a_\nu a_{\nu'}^\dagger|\psi\rangle}_{=2\pi\delta(\nu-\nu')} \\ &= \int \frac{d\nu}{2\pi} \beta_{\omega\nu}^* \beta_{\omega'\nu} \neq 0. \end{aligned} \quad (4.13)$$

This phenomenon already happens in flat spacetime when one considers an accelerated observer. In this case it is called the *Unruh effect*.

4.2 Hawking radiation

Let us apply what we learned above about QFT in curved space to black holes. We will present a simplified, “baby” derivation of Hawking radiation in the background of a Schwarzschild black hole. This reproduces the discussion in [7], see also [8].

We start from the Schwarzschild metric

$$ds^2 = -\left(1 - \frac{r_s}{r}\right) dt^2 + \frac{dr^2}{\left(1 - \frac{r_s}{r}\right)} + r^2 d\Omega_2, \quad (4.14)$$

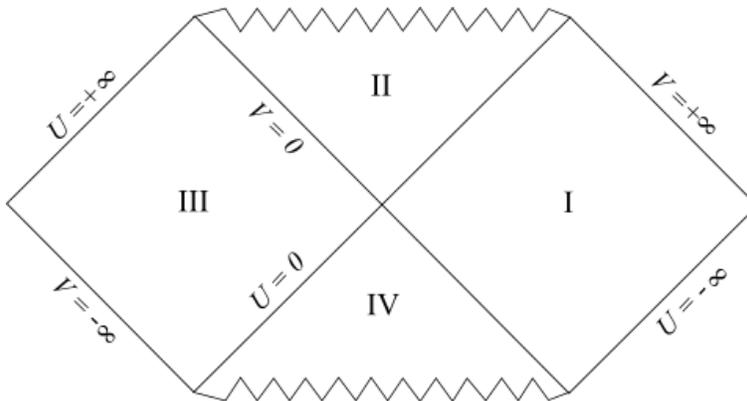
where $r_s = 2GM$ is the Schwarzschild radius and $d\Omega_2 = d\theta^2 + \sin^2\theta d\phi^2$ is the unit metric on the two-sphere. Consider an observer freely falling through the (future) event horizon. This observer will naturally use a set of coordinates that is well defined across the horizon, such as the null Kruskal coordinates given by

$$UV = r_s(r_s - r) e^{r/r_s}, \quad \frac{U}{V} = -e^{-t/r_s}. \quad (4.15)$$

In these coordinates the metric reads

$$ds^2 = -\frac{4r_s}{r} e^{-r/r_s} dUdV + r^2 d\Omega_2. \quad (4.16)$$

The position of the horizon, $r = r_s$, corresponds to $U = 0$; we see that the metric is perfectly regular there. The curvature singularity is at $UV = r_s^2$. Outgoing null geodesics correspond to $U = \text{const}$, while ingoing null geodesics correspond to $V = \text{const}$. The original t, r coordinates only cover the quadrant I. We can restrict our attention to quadrants I and II in the picture; these define the region of spacetime relevant when the black hole is formed from a collapsing body.



The trajectory of the infalling observer is described by $V \sim \text{const}$ while U goes to zero linearly in their proper time τ . An asymptotic observer sees the Minkowski metric around them and will naturally use the t, r coordinates; in the t coordinate, the infalling observer takes an infinitely long time to reach the horizon. The infalling observer proper time τ is related to the time t of the asymptotic observer as

$$d\tau \propto e^{-t/r_s} dt. \quad (4.17)$$

Hence there is an exponential redshift factor between $d\tau$ and dt : a short proper time for the infalling observer is perceived as a long time for the asymptotic one. The relation between these two times is at the origin of Hawking radiation. Recall what we saw in Section 4.1: if two observers use different timelike Killing vectors to define the frequency expansion of a quantum field, an empty vacuum for one of them will be full of particles for the other.

We would like to argue that the infalling observer indeed sees an empty vacuum. This is because of the adiabatic principle. This principle says that if the parameters in the Hamiltonian of a quantum system change slowly compared to the spacing between the energy levels, then the probability of an excitation is exponentially small. In other words, you will stay in the ground state with very high probability. For our infalling observer, the geometry is changing adiabatically on a time characteristic scale r_s^{-1} , while any mode that the asymptotic observer may detect as Hawking quanta are at very high frequency ν for the infalling observer (since they are exponentially blueshifted if we trace them back from infinity to near the horizon). So to a very high accuracy $e^{-\mathcal{O}(\nu r_s)}$ these modes will not be excited.

Let us make this quantitative by considering the very simplified setup of a massless scalar field φ in the $1+1$ Schwarzschild geometry. This is obtained from the four-dimensional Schwarzschild geometry by ignoring the angular directions. We can imagine we are looking

at the black hole long after it has formed, so that the geometry is static. The metric can be written as

$$\begin{aligned}
ds^2 &= - \left(1 - \frac{r_s}{r}\right) dt^2 + \frac{dr^2}{\left(1 - \frac{r_s}{r}\right)} \\
&= - \left(1 - \frac{r_s}{r}\right) du dv \\
&= - \frac{4r_s}{r} e^{-r/r_s} dU dV,
\end{aligned} \tag{4.18}$$

where the expression in the last line is appropriate for the infalling observer using the Kruskal coordinates U, V that are well defined at the horizon, while the second line uses the Eddington-Finkelstein coordinates

$$u = t - r_* = -2r_s \log(-U/r_s), \quad v = t + r_* = 2r_s \log(V/r_s), \tag{4.19}$$

where $r_* = r + r_s \log(r - r_s)$ is the tortoise radial coordinate. The u, v coordinates are appropriate null coordinates for the asymptotic observer, as they are defined in the first quadrant and are linear in the Minkowski time t .

The two-dimensional Klein-Gordon equation takes the same form in the two coordinate systems,

$$\partial_u \partial_v \varphi = \partial_U \partial_V \varphi = 0, \tag{4.20}$$

leading to ingoing (left-moving) and outgoing (right-moving) wave solutions for both observers. The ingoing solutions are functions of V (or v), while the outgoing ones are function of U (or u). Let us focus on the outgoing part of the field, φ_R (where ‘‘R’’ stands for ‘‘right moving’’, as appropriate for propagation in 1+1 dimensions). The infalling observer uses an expansion in terms of $e^{-i\nu U}$ modes with frequency ν as

$$\varphi_R = \int_0^\infty \frac{d\nu}{2\pi\sqrt{2\nu}} (a_\nu e^{-i\nu U} + a_\nu^\dagger e^{i\nu U}), \tag{4.21}$$

with $[a_\nu, a_{\nu'}^\dagger] = 2\pi\delta(\nu - \nu')$, $[a_\nu, a_{\nu'}] = 0$, $[a_\nu^\dagger, a_{\nu'}^\dagger] = 0$, while the asymptotic observer expands the field in terms of ω -frequency $e^{-i\omega u}$ modes as

$$\varphi_R = \int_0^\infty \frac{d\omega}{2\pi\sqrt{2\omega}} (b_\omega e^{-i\omega u} + b_\omega^\dagger e^{i\omega u}), \tag{4.22}$$

with $[b_\omega, b_{\omega'}^\dagger] = 2\pi\delta(\omega - \omega')$, etc. Taking a Fourier transform, we can express the b_ω operators

in term of a_ν, a_ν^\dagger ,

$$\begin{aligned}
b_\omega &= \sqrt{2\omega} \int du e^{i\omega u} \varphi_R \\
&= \sqrt{2\omega} \int du e^{i\omega u} \int_0^\infty \frac{d\nu}{2\pi\sqrt{2\nu}} (a_\nu e^{-i\nu U} + a_\nu^\dagger e^{i\nu U}) \\
&= \int_0^\infty \frac{d\nu}{2\pi} (\alpha_{\omega\nu} a_\nu + \beta_{\omega\nu} a_\nu^\dagger), \tag{4.23}
\end{aligned}$$

with the Bogoliubov coefficients α, β being given by

$$\begin{aligned}
\alpha_{\omega\nu} &= \sqrt{\frac{\omega}{\nu}} \int du e^{i\omega u - i\nu U}, \\
\beta_{\omega\nu} &= \sqrt{\frac{\omega}{\nu}} \int du e^{i\omega u + i\nu U}. \tag{4.24}
\end{aligned}$$

Working out the integrals, these take the form

$$\begin{aligned}
\alpha_{\omega\nu} &= 2r_s \sqrt{\frac{\omega}{\nu}} (r_s \nu)^{2ir_s \omega} e^{\pi r_s \omega} \Gamma(-2ir_s \omega), \\
\beta_{\omega\nu} &= 2r_s \sqrt{\frac{\omega}{\nu}} (r_s \nu)^{2ir_s \omega} e^{-\pi r_s \omega} \Gamma(-2ir_s \omega), \tag{4.25}
\end{aligned}$$

where Γ is the Euler gamma function.

Let us prove the second in (4.25) (the proof of the first being analogous). We start from

$$u = -2r_s \log\left(-\frac{U}{r_s}\right) \quad \Rightarrow \quad du = -2r_s \frac{dU}{U}. \tag{4.26}$$

Hence

$$\int_{-\infty}^{+\infty} du e^{i\omega u + i\nu U} = 2 \int_{-\infty}^0 dU e^{i\nu U} \left(-\frac{U}{r_s}\right)^{-2ir_s \omega - 1}. \tag{4.27}$$

Let us set

$$U = -\frac{\rho}{\nu}, \quad z = -2ir_s \omega. \tag{4.28}$$

Then the expression above becomes

$$\int_{-\infty}^{+\infty} du e^{i\omega u + i\nu U} = 2r_s (r_s \nu)^{-z} \int_0^\infty d\rho e^{-i\rho} \rho^{z-1} = 2r_s (r_s \nu)^{-z} e^{-\pi iz} \Gamma(z). \tag{4.29}$$

Using this in the definition (4.24) of $\beta_{\omega\nu}$ we arrive at the second in (4.25). ■

We now have all the elements to see the Hawking radiation. Using the adiabatic principle, we argue that the infalling observer sees the a -modes as empty, $a_\nu |\psi\rangle = 0$, where $|\psi\rangle$ is the

state in which the field is. Then the b -modes will not be empty. We can compute the occupation number for these outgoing modes. We have

$$\langle \psi | b_{\omega}^{\dagger} b_{\omega'} | \psi \rangle = \int_0^{\infty} \frac{d\nu}{2\pi} \int_0^{\infty} \frac{d\nu'}{2\pi} \beta_{\omega\nu}^* \beta_{\omega'\nu'} \underbrace{\langle \psi | a_{\nu} a_{\nu'}^{\dagger} | \psi \rangle}_{=2\pi\delta(\nu-\nu')} = \int \frac{d\nu}{2\pi} \beta_{\omega\nu}^* \beta_{\omega'\nu} \quad (4.30)$$

as already seen (recall (4.13)). Plugging the expression (4.25) for the β Bogoliubov coefficients into (4.30), we get

$$\langle \psi | b_{\omega}^{\dagger} b_{\omega'} | \psi \rangle = \sqrt{\omega\omega'} 4r_s^2 (r_s)^{2ir_s(\omega'-\omega)} e^{-\pi r_s(\omega+\omega')} \Gamma^*(-2ir_s\omega) \Gamma(-2ir_s\omega') I, \quad (4.31)$$

where

$$I = \int_0^{\infty} \frac{d\nu}{2\pi\nu} \nu^{2ir_s(\omega'-\omega)} \underbrace{=}_{\nu=e^{\rho/(2r_s)}} \frac{1}{2r_s} \int_{-\infty}^{\infty} \frac{d\rho}{2\pi} e^{i\rho(\omega'-\omega)} = \frac{1}{2r_s} \delta(\omega - \omega'). \quad (4.32)$$

therefore

$$\langle \psi | b_{\omega}^{\dagger} b_{\omega'} | \psi \rangle = 2r_s\omega e^{-2\pi r_s\omega} |\Gamma(-2ir_s\omega)|^2 \delta(\omega - \omega'). \quad (4.33)$$

For real y the Gamma function satisfies

$$|\Gamma(iy)|^2 = \frac{\pi}{y \sinh(\pi y)}. \quad (4.34)$$

Using this in (4.33) for $y = -2r_s\omega$, it is easy to see that our formula becomes

$$\langle \psi | b_{\omega}^{\dagger} b_{\omega'} | \psi \rangle = \frac{2\pi \delta(\omega - \omega')}{e^{4\pi r_s\omega} - 1}. \quad (4.35)$$

Through the manipulations above, we have showed that

$$\begin{aligned} \langle \psi | b_{\omega}^{\dagger} b_{\omega'} | \psi \rangle &= \int \frac{d\nu}{2\pi} \beta_{\omega\nu}^* \beta_{\omega'\nu} \\ &= \frac{2\pi \delta(\omega - \omega')}{e^{\hbar\omega/T_H} - 1}, \quad \text{with } T_H = \frac{\hbar}{4\pi r_s} = \frac{\hbar\kappa}{2\pi}, \end{aligned} \quad (4.36)$$

where we have used that for the Schwarzschild black hole the surface gravity is $\kappa = \frac{1}{4GM} = \frac{1}{2r_s}$. Hence the spectrum of the outgoing modes is a thermal blackbody spectrum, with temperature T_H .

This computation fixes the numerical coefficient α introduced under eq. (3.6) to $\alpha = 4$. Hence the final formulae for the *Hawking temperature* and *Bekenstein-Hawking entropy* read

$$T = \frac{\kappa}{2\pi}, \quad S = \frac{A}{4}. \quad (4.37)$$

One can show that these expressions are still valid when one adds angular momentum and charge to the black hole. They also hold in different spacetime dimensions. So they are very universal.

Pair production. Since the U, V coordinates are well defined both in quadrant I and II, the a -expansion (4.21) of φ_R is valid both outside and inside the horizon. On the other hand, the u, v coordinates only cover region I, so the b -expansion (4.22) is only valid there. This implies that while the expression (4.23) for b_ω in terms of a, a^\dagger is complete, the inverse relation expressing a, a^\dagger also involves some other operators, \tilde{b}_ω , whose modes have support only in region II inside the horizon. One has

$$a_\nu = \int_0^\infty \frac{d\omega}{2\pi} \left(\alpha_{\omega\nu}^* b_\omega - \beta_{\omega\nu}^* b_\omega^\dagger + \tilde{\alpha}_{\omega\nu}^* \tilde{b}_\omega - \tilde{\beta}_{\omega\nu}^* \tilde{b}_\omega^\dagger \right). \quad (4.38)$$

One can see that $[H, b_\omega^\dagger] = \omega b_\omega^\dagger$ and $[H, \tilde{b}_\omega^\dagger] = -\omega \tilde{b}_\omega^\dagger$, hence while the creation operator b_ω^\dagger raises the energy by ω , the creation operator \tilde{b}_ω^\dagger lowers the energy by ω . The modes created by \tilde{b}_ω^\dagger are in fact necessary for energy conservation: every time a particle with positive energy ω is created and propagates away from the black hole horizon, a particle with negative energy $-\omega$ is also created, and falls into the horizon. These particles with opposite energy are entangled; so there is a large entanglement between the radiation propagating outside the horizon and the inside. The resulting state for our quantum field in the black hole background is described by the repeated action of $\tilde{b}_\omega^\dagger b_\omega^\dagger$ on the vacuum $|0\rangle_{b, \tilde{b}}$. Its actual form is

$$|0\rangle_a \propto \exp \left(\int_0^\infty \frac{d\omega}{2\pi} e^{-\omega/2T_H} b_\omega^\dagger \tilde{b}_\omega^\dagger \right) |0\rangle_{b, \tilde{b}}, \quad (4.39)$$

which is the form of a squeezed state.

So we can interpret the Hawking emission process as arising from particle pair creation close to the horizon, with a negative energy particle falling into the black hole and a positive energy particle escaping to infinity. One may be surprised by the appearance of propagating negative energy modes. However, one should recall that here the energy is the conserved charge associated with a Killing vector that generates time translations far away from the horizon, let's say ∂_t . This vector is timelike outside the horizon, but becomes spacelike inside the horizon; the charge of a spacelike Killing vector is momentum, and this can be either positive, or negative so there is no worry. We see that since Hawking radiation needs a timelike Killing vector becoming spacelike, it can only happen in the vicinity of a horizon.

4.3 Further remarks

Some numbers. Let us quantify the Hawking temperature. For a Schwarzschild black hole, $\kappa = \frac{c^4}{4GM}$ and therefore

$$T = \frac{\hbar c^3}{8\pi G k_B M} \simeq (6 \cdot 10^{-8} \text{ K}) \frac{M_{\text{Sun}}}{M}. \quad (4.40)$$

So a black hole of one solar mass has a tiny Hawking temperature, and would absorb far more cosmic microwave background radiation than it emits. Observing Hawking radiation emitted by solar mass black holes is thus hopeless. In order to be in equilibrium with the cosmic microwave background at 2.7 K, the black hole should have a mass of $4.5 \cdot 10^{22}$ kg, that is roughly the mass of the Moon. Smaller primordial black holes would emit more than they absorb and hence evaporate.

Negative heat capacity. Since the temperature is inversely proportional to the mass, $T = \frac{1}{8\pi M}$, the Schwarzschild black hole gets hotter as long as it loses mass via evaporation. It has *negative heat capacity*

$$C = \frac{dM}{dT} = -\frac{1}{8\pi T^2} < 0. \quad (4.41)$$

This signals a thermodynamical instability: if we start from a black hole at equilibrium with a thermal bath (i.e. they have the same temperature), then emission will prevail over absorption. The evaporation becomes faster and faster as long as it goes on, until reaching a final explosion.

Information paradox. Black hole evaporation leads to a serious problem with unitarity. Consider a black hole that forms from collapsing matter and then evaporates away completely, leaving just thermal radiation. It should be in principle possible to arrange that the collapsing matter is in a definite quantum state $|\psi\rangle$; the associated density matrix would be the one of a pure state, namely just the projector $\rho = |\psi\rangle\langle\psi|$. When the black hole is formed, the Hilbert space \mathcal{H} naturally splits into the tensor product of a Hilbert space of states with support in the interior of the black hole, and a space of states with support outside the horizon, $\mathcal{H} = \mathcal{H}_{\text{in}} \otimes \mathcal{H}_{\text{out}}$. An outside observer does not have access to \mathcal{H}_{in} , so their description of the black hole state is necessarily incomplete: they will describe the state outside the horizon by means of a reduced density matrix, obtained by tracing over \mathcal{H}_{in} : $\rho_{\text{out}} = \text{Tr}_{\text{in}} \rho$. Since it is described by a non-trivial density matrix, the outside state is mixed. This is consistent with the fact that it contains thermal radiation, and so far there is no issue, as the external state is entangled with the interior; the reduced density matrix ρ_{out} is just a way in which the outside observer expresses their ignorance about part of the system. However if we assume that after the black hole has completely evaporated nothing is left in the interior, the exterior reduced density matrix ρ_{out} will describe the full state, which is therefore a mixed state. But evolution from a pure state into a mixed state is forbidden by unitarity of quantum mechanics.

This is, in extreme synthesis, the black hole information paradox. It is important to

emphasize the difference with thermal radiation produced in ordinary physical processes, which do not violate unitarity. If a book is burned, thermal radiation is produced, however the process is unitary and in principle one could reconstruct all the information contained in the book by studying very carefully the radiation and the ashes that are left over. The early radiation is entangled with excitations inside the burning body; however the excitations inside the burning body can still transmit information to the radiation emitted later on, which will thus contain non-trivial information. By contrast, if the book is thrown into a black hole, the information appears to be really lost once evaporation is completed, because the final radiation is *exactly* thermal. Indeed the internal excitations are shielded by the horizon, and by causality they cannot influence the later outgoing radiation.

After more than forty years since it was formulated by Hawking, the black hole information paradox is still an open problem and an active area of research. We will not discuss it further in these lectures, see e.g. [7, 9] for an introduction and possible resolutions.

5 Euclidean Quantum Gravity

An entirely different approach to black hole thermodynamics is given by Euclidean Quantum Gravity. This approach was pioneered by Hawking, Gibbons, and others in the Seventies.

5.1 QFT at finite temperature

We saw that a quantum field in the black hole background emits thermal radiation. So it seems a good idea to study QFT at finite temperature in the same background. This should be seen as a low-energy limit of the full Quantum Gravity, such that the gravitational degrees of freedom are not excited (this makes sense because in dimension $d > 2$, the gravitational interaction, being controlled by the dimensionful coupling constant $G = (M_P)^{2-d}$, is technically irrelevant, so it is not important at low energy).

Canonical ensemble

QFT at finite temperature is the same as QFT with an imaginary time periodicity

$$t \sim t + i\beta, \quad \text{where } \beta = 1/T. \quad (5.1)$$

Let us recall why this is true by considering a thermal Green's function, for instance the two-point function for some operator $O(t, x)$; here x denotes just the spatial coordinates.

We assume time invariance, so that we can say we are studying an equilibrium state at a certain temperature T . In other words, we set ourselves in the *canonical ensemble*. Given two operators $O(t, x)$ and $O(t', x')$, the time dependence of the Green's function is just via the difference $t - t'$, and by a time translation we can choose $t' = 0$. The Green's function thus takes the form $G_\beta = G_\beta(t; x, x')$, where $\beta = 1/T$. Starting from the Hamiltonian H , generating translations along t , one introduces the canonical density matrix

$$\rho = e^{-\beta H}, \quad (5.2)$$

and the *canonical partition function*,

$$Z(\beta) = \text{Tr} e^{-\beta H}, \quad (5.3)$$

which is the trace of the density matrix over the Hilbert space of the theory. The thermal average of any operator O is given by $\langle O \rangle_\beta = Z(\beta)^{-1} \text{Tr} (e^{-\beta H} O)$. In particular, the Green's function is defined as

$$G_\beta(t, x, x') = Z^{-1} \text{Tr} e^{-\beta H} O(t, x) O(0, x'), \quad (5.4)$$

where we are assuming $t > 0$, so that the operators are time-ordered. Recall that the time evolution in the Heisenberg picture is

$$O(t + \Delta t, x) = e^{i\Delta t H} O(t, x) e^{-i\Delta t H}. \quad (5.5)$$

If we allow ourselves to analytically continue this by choosing an imaginary time interval $\Delta t = i\beta$, we get

$$O(t + i\beta, x) = e^{-\beta H} O(t, x) e^{\beta H}. \quad (5.6)$$

Using this in our Green's function (5.4) we obtain

$$\begin{aligned} G_\beta(t, x, x') &= Z^{-1} \text{Tr} O(t + i\beta, x) e^{-\beta H} O(0, x') \\ &= Z^{-1} \text{Tr} e^{-\beta H} O(0, x') O(t + i\beta, x) \\ &= (-1)^F Z^{-1} \text{Tr} e^{-\beta H} O(t + i\beta, x) O(0, x') \\ &= (-1)^F G_\beta(t + i\beta, x, x'), \end{aligned} \quad (5.7)$$

where $F = 0$ if the operators are bosonic, while $F = 1$ if they are fermionic. In the second line we used cyclicity of the trace and in the third the fact that the operators at distinct points commute if they are bosonic, and anticommute if they are fermionic. Thus we have found that the Green's function is (anti-)periodic, with an *imaginary* time period $i\beta$.

One can see that all other thermal correlation functions satisfy the same periodicity property. The converse is also true: if all Green's functions are periodic (anti-periodic in the case of fermionic operators) with an imaginary time period $i\beta$, then they must have been computed in the canonical ensemble at temperature $T = 1/\beta$. This is called the KMS condition [Kubo, Martin, Schwinger].

Since we have to deal with an imaginary time period, it is convenient to Wick-rotate to the Euclidean time $\tau = it$. The Euclidean Green's functions $G^E(\tau, x) = G(t = -i\tau, x)$ satisfy

$$G_\beta^E(\tau; x, x') = (-1)^F G_\beta^E(\tau - \beta; x, x'), \quad (5.8)$$

namely they are periodic (if bosonic) or anti-periodic (if fermionic) in Euclidean time, with period β .

We conclude that QFT at temperature T is equivalent to QFT in periodic Euclidean time, with period $\beta = 1/T$. Usually it is convenient to make all the computations in Euclidean signature and analytically continue back to the Lorentzian spacetime at the end.

Path integral representation

We will find it useful to take the path integral point of view. Recall that the path integral computes the amplitude to go from an initial field configuration φ_1 at time τ_1 to a final configuration φ_2 at time τ_2 as

$$\langle \varphi_2, \tau_2 | \varphi_1, \tau_1 \rangle = \int \mathcal{D}\varphi e^{-I_E[\varphi]}, \quad (5.9)$$

where the path integration is over all configurations of φ that interpolate between φ_1 at time τ_1 and φ_2 at time τ_2 . But this amplitude is the same as

$$\langle \varphi_2, \tau_2 | \varphi_1, \tau_1 \rangle = \langle \varphi_2 | e^{-(\tau_2 - \tau_1)H} | \varphi_1 \rangle, \quad (5.10)$$

where the relation between the Euclidean action I_E and the Hamiltonian H is $I_E[\varphi] = \int d\tau (-i\Pi\dot{\varphi} + H)$, where $\Pi = \frac{\delta I_E}{\delta \dot{\varphi}}$ is the canonical momentum. Taking $\tau_2 - \tau_1 = \beta$, $\varphi_2 = \varphi_1$ and then summing over all boundary conditions φ_1 , we obtain the path integral representation of the canonical partition function,

$$Z(\beta) = \text{Tr} e^{-\beta H} = \int \mathcal{D}\varphi e^{-I_E[\varphi]}, \quad (5.11)$$

where the integral is performed over fields that are periodic (if bosonic) or antiperiodic (if fermionic) in Euclidean time, with period β . The thermal correlation functions can be obtained by including operator insertions into the path integral.

Grand-canonical ensemble

The discussion above can be extended to the grand-canonical ensemble, where in addition to the temperature we specify “chemical potentials” for one or more conserved quantities. We will be interested in the case where the conserved quantities are an angular momentum J , generating rotations along an angle ϕ , and a U(1) charge Q . It is assumed that H, J and Q are commuting operators. The corresponding grand-canonical density matrix is

$$\rho = e^{-\beta(H - \Omega J - \Phi Q)}, \quad (5.12)$$

where Ω is the angular potential for the rotations generated by J , and Φ is the potential for the U(1) transformations generated by Q . The grand-canonical partition function is

$$Z(\beta, \Omega, \Phi) = \text{Tr} e^{-\beta(H - \Omega J - \Phi Q)}. \quad (5.13)$$

Generalising the argument above, one can show (do this as an **exercise**) that the Green’s functions for an operator with given U(1) charge q satisfy⁹

$$G_{\beta, \Omega, \Phi}(t, \phi, x, x') = (-1)^F e^{q\beta\Phi} G_{\beta, \Omega, \Phi}(t + i\beta, \phi + i\beta\Omega, x, x'), \quad (5.14)$$

namely they are periodic in imaginary time, but with an extra shift in the angular direction and with a specific rescaling factor related to the charge.

Therefore QFT at temperature T , angular potential Ω and electric potential Φ is equivalent to QFT in a background having the coordinate identification

$$(t, \phi) \sim (t + i\beta, \phi + i\beta\Omega), \quad (5.15)$$

and with the correlation functions being identified up to a global U(1) transformation with imaginary parameter $= -i\beta\Phi$.

A somewhat simpler picture is obtained by introducing the new coordinates

$$\hat{\phi} = \phi - \Omega t, \quad \hat{t} = t, \quad (5.16)$$

so that the identification (5.15) only involves a shift of the new time coordinate,

$$(\hat{t}, \hat{\phi}) \sim (\hat{t} + i\beta, \hat{\phi}). \quad (5.17)$$

Notice that if $H = i\partial_t$ generates translations along t and $J = -i\partial_\phi$ generates translations along ϕ , the combined operator

$$H - \Omega J = i(\partial_t + \Omega\partial_\phi) = i\partial_{\hat{t}} \quad (5.18)$$

⁹Here x denotes the spatial coordinates different from t, ϕ , namely $x = \{r, \theta\}$.

generates precisely the translations along \hat{t} . Now we can introduce the Euclidean time $\hat{\tau} = i\hat{t}$ and obtain

$$(\hat{\tau}, \hat{\phi}) \sim (\hat{\tau} - \beta, \hat{\phi}), \quad (5.19)$$

so the coordinates have standard identifications when we go once around the circle of length β parameterized by $\hat{\tau}$. Where has the angular potential Ω gone? It now appears in the spacetime metric. If the original metric was of the form

$$ds^2 = -f(r, \theta) dt^2 + h(r, \theta) d\phi^2 + \text{rest}, \quad (5.20)$$

after the change of coordinates it reads

$$\begin{aligned} ds^2 &= -f(r, \theta) d\hat{t}^2 + h(r, \theta) (d\hat{\phi} + \Omega d\hat{t})^2 + \text{rest} \\ &= f(r, \theta) d\hat{\tau}^2 + h(r, \theta) (d\hat{\phi} - i\Omega d\hat{\tau})^2 + \text{rest}, \end{aligned} \quad (5.21)$$

so after the Wick rotation it has some imaginary components. One can also undo the twisted identification of the fields by the $U(1)$ transformation by gauging it and performing a gauge transformation with parameter $\lambda = -i\Phi\hat{\tau}$. Indeed, the gauge-transformed fields are related to the old ones as $\varphi^{\text{new}} = e^{iq\lambda}\varphi^{\text{old}} = e^{q\Phi\hat{\tau}}\varphi^{\text{old}}$; so when we go around the Euclidean time circle parameterized by $\hat{\tau}$ the old fields satisfy $\varphi^{\text{old}}(\hat{\tau} - \beta) \sim e^{q\beta\Phi}\varphi^{\text{old}}(\hat{\tau})$, but the new ones are periodic, $\varphi^{\text{new}}(\hat{\tau} - \beta) \sim \varphi^{\text{new}}(\hat{\tau})$. This gauge transformation introduces a background gauge field

$$A = \Phi d\hat{t} = -i\Phi d\hat{\tau} \quad (5.22)$$

minimally coupled to the dynamical fields in the QFT. Indeed, $A^{\text{new}} = A^{\text{old}} + d\lambda = 0 + d\lambda = -i\Phi d\hat{\tau}$. So we have traded the twisted identification for the background field.

Treating $\hat{H} = H - \Omega J - \Phi Q$ as the actual Hamiltonian, one can derive the corresponding Lagrangian entering in the path integral representation of the grand-canonical partition function.¹⁰ One finds

$$Z(\beta, \Omega, \Phi) = \text{Tr} e^{-\beta(H - \Omega J - \Phi Q)} = \int \mathcal{D}\varphi e^{-I_E[\varphi, g, A]}, \quad (5.23)$$

where the field theory is now defined on a complex background metric of the form (5.21), and is minimally coupled to the background gauge field (5.22). The fields are taken periodic in the Euclidean time circle of length β parameterized by $\hat{\tau}$.

¹⁰For the effect of the ΦQ term see for instance Section 3.2 of M. Le Bellac, *Thermal field theory*, CUP, 1996.

5.2 Hawking temperature from regularity of Euclidean geometry

Suppose we want to compute thermal correlation functions in the background of a Schwarzschild black hole. As we recalled above, this can be done by considering a periodic Euclidean time. Let's go for it.

With $t = -i\tau$, the Schwarzschild metric (4.14) becomes

$$ds^2 = \left(1 - \frac{r_s}{r}\right) d\tau^2 + \frac{dr^2}{\left(1 - \frac{r_s}{r}\right)} + r^2 d\Omega_2. \quad (5.24)$$

Let us study this metric. As $r \rightarrow \infty$, the metric is the flat one on $S^1 \times \mathbb{R}$. Moving towards lower values of r , nothing special happens until we reach $r \rightarrow r_s$, where $g_{\tau\tau} \rightarrow 0$ and $g_{rr} \rightarrow \infty$. For $r < r_s$ instead the metric has mixed $(--++)$ signature, and does not describe the same space. So we should think of the region connected with infinity as being described by $r \geq r_s$, with the space ending at $r = r_s$. In this way the curvature singularity in $r = 0$ is excluded from the space of interest.

Let us examine more closely what happens as r approaches r_s . We introduce a new coordinate ρ as

$$r = r_s + \frac{\rho^2}{4r_s}, \quad \text{with } \rho \ll r_s. \quad (5.25)$$

Using $dr = \frac{\rho}{2r_s} d\rho$ and $1 - \frac{r_s}{r} = \frac{\rho^2}{4r_s^2} + \dots$, the metric reads at leading order near r_s

$$ds^2 = \rho^2 \frac{d\tau^2}{4r_s^2} + d\rho^2 + r_s^2 d\Omega_2 + \dots \quad (5.26)$$

This is the metric on $\mathbb{R}^2 \times S^2$, where S^2 has radius r_s and \mathbb{R}^2 is parameterized in polar coordinates. Therefore $\frac{\tau}{2r_s}$ plays the role of an angular coordinate. We really obtain \mathbb{R}^2 if this angular coordinate is identified with period 2π , otherwise we have a conical singularity in the $\rho - \tau$ plane at $\rho = 0$.¹¹ So we must take

$$\tau \sim \tau + \beta, \quad \text{with} \quad \beta = 4\pi r_s = \frac{2\pi}{\kappa} = \frac{1}{T_H}. \quad (5.27)$$

So we have found that regularity of the Euclidean Schwarzschild metric requires the Euclidean time to be periodic with period given by the inverse Hawking temperature!

¹¹If we identify the angular coordinate with a period $2\pi - \Theta$, then the space is a cone, with deficit angle Θ . This can be visualized by embedding our surface in \mathbb{R}^3 . The tip of the cone is singular as the curvature is a delta function peaked there. One way to see this is to smoothen out the cone by a small cap and then shrink it off: the curvature will be more and more peaked around the tip until when it becomes a delta function in the limit. We do not allow for a conical singularity as it does not solve the vacuum Einstein equation.

The geometry described by the Riemannian metric (5.24), and with the coordinates satisfying $r_s \leq r < \infty$, $\tau \sim \tau + \beta$, $0 \leq \theta \leq \pi$, $\phi \sim \phi + 2\pi$, is perfectly regular. It is called the *Euclidean section* of the Schwarzschild solution. In particular, the two-dimensional hypersurface at fixed θ, ϕ , parameterized by r, τ , asymptotically looks like a cylinder, while as $r \rightarrow r_s$ caps off smoothly; so it has the shape of a cigar.

All Green's functions of a quantum field on this background have a periodicity in τ of T_H^{-1} . The KMS condition then implies we are in the canonical ensemble at the Hawking temperature T_H . So the canonical partition function reads

$$Z(\beta) = \text{Tr} e^{-\beta H}, \quad (5.28)$$

and we can define the Green's functions for our quantum field by including the corresponding operator in the trace. Therefore we are describing a gas at temperature T_H in equilibrium with the black hole. By the zeroth law of thermodynamics, it follows that the black hole itself has the temperature T_H , and since we are at equilibrium it must be able to emit as much as it absorbs. This equilibrium state is called the Hartle-Hawking state.

We can also take the path integral point of view and state that the canonical partition function in the black hole background is computed by an Euclidean path integral with fields periodic in the Euclidean time, with period $\beta = T_H^{-1}$.

5.3 Regularity of Kerr-Newman and grand-canonical ensemble

We analyze the Euclidean section of the Kerr-Newman solution. We take $P = 0$ for simplicity. Consider first the metric (2.61), where it is convenient to use $\Delta = (r - r_+)(r - r_-)$, without substituting the parameters M, a, Q in r_{\pm} . Redefining the radial coordinate as

$$r = r_+ + \frac{\rho^2}{r_+}, \quad (5.29)$$

one can show that close to $\rho = 0$ the metric takes the form

$$ds^2 = g_{\rho\rho} (d\rho^2 - \rho^2 \kappa^2 dt^2) + g_{\theta\theta} d\theta^2 + g_{\phi\phi} (d\phi - \Omega dt - \omega \rho^2 dt)^2, \quad (5.30)$$

where

$$\kappa = \frac{r_+ - r_-}{2(a^2 + r_+^2)}, \quad \Omega = \frac{a}{r_+^2 + a^2} \quad (5.31)$$

are the same as the surface gravity (2.76) and the angular velocity (2.73) of the horizon, while $g_{\rho\rho}, g_{\theta\theta}, g_{\phi\phi}, \omega$ have an expansion in powers of ρ whose leading-order, $\mathcal{O}(\rho^0)$, term is a non-vanishing function of the coordinate θ and of the parameters a, r_{\pm} , (in order to fix the $\rho^2 d\tau^2$ terms in (5.30) one needs to include the $\mathcal{O}(\rho^2)$ term in $g_{\phi\phi}$ and the $\mathcal{O}(\rho^0)$ term in ω).

In this rotating solution, the vector whose norm goes to zero as $\rho \rightarrow 0$ is $\xi = \partial_t + \Omega \partial_\phi$; this defines the direction that should be identified as the Euclidean time. In order to see the correct regularity condition to be imposed, perform the coordinate transformation

$$\hat{\phi} = \phi - \Omega t, \quad \hat{t} = t, \quad (5.32)$$

so that $\xi = \partial_{\hat{t}}$ and the metric reads

$$ds^2 = g_{\rho\rho} (d\rho^2 - \rho^2 \kappa^2 d\hat{t}^2) + g_{\theta\theta} d\theta^2 + g_{\phi\phi} \left(d\hat{\phi} - \omega \rho^2 d\hat{t} \right)^2. \quad (5.33)$$

Now we can Wick rotate $\hat{t} = -i\hat{\tau}$. We see that the correct regularity condition for the two-dimensional cigar geometry parameterized by $(\rho, \hat{\tau})$ to close off smoothly is that

$$(\hat{\tau}, \hat{\phi}) \sim (\hat{\tau} + \beta, \hat{\phi}), \quad (5.34)$$

with $\beta = 2\pi/\kappa = T_H^{-1}$. In the original coordinates, this identification is equivalent to $(t, \phi) \sim (t - i\beta, \phi - i\beta\Omega)$.

We should also study the gauge field. At leading order near to $\rho \rightarrow 0$, the gauge field (2.62) (with $P = 0$) reads

$$A = -\Phi d\hat{t} + \frac{ar_+ Q \sin^2 \theta}{r_+^2 + a^2 \cos^2 \theta} d\hat{\phi} + \mathcal{O}(\rho^2), \quad (5.35)$$

where

$$\Phi = \frac{Q r_+}{r_+^2 + a^2} \quad (5.36)$$

is the same as the electric potential (2.80) of the horizon.

This gauge field is singular in $\rho = 0$; one way to see it is that the norm of $A_\mu A^\mu$ diverges as $\rho \rightarrow 0$, as $g^{\hat{t}\hat{t}}$ goes to infinity. A regular gauge field is obtained by making the gauge shift

$$A \rightarrow \hat{A} = A + \Phi d\hat{t}, \quad (5.37)$$

which removes the problematic $d\hat{t}$ term.

We have thus identified a regular section of the solution. Note that both the metric and the gauge field are complex. We could obtain a real, positive definite metric by analytically continuing $a = i\hat{a}$. One could do all the computations in this real Euclidean section and then analytically continue the parameter a back to the original value.

Let us go and see what happens near to infinity. In the coordinates $\hat{\tau}, \hat{\phi}$, the solution at large r is

$$ds^2 \rightarrow dr^2 + r^2 \left(d\hat{\tau}^2 + d\theta^2 + \sin^2 \theta \left(d\hat{\phi} - i\Omega d\hat{\tau} \right)^2 \right), \quad (5.38)$$

$$A \rightarrow -i\Phi d\hat{\tau}. \quad (5.39)$$

The asymptotic observer thus is co-rotating with the hole at the same angular velocity Ω and is immersed in the same electric potential Φ as the one of the hole. The observer at infinity is thus at equilibrium with the hole in the grand-canonical ensemble.

We conclude that regularity of the Kerr-Newman Euclidean solution implies that QFT in this background is at finite temperature $T = \frac{\kappa}{2\pi}$, finite angular potential $\Omega = \Omega_H$ and electric potential $\Phi = \Phi_H$.

Exercise. Check the steps above.

5.4 The gravitational path integral

So far we have been playing with QFT in a curved but fixed background. Now we want to be more ambitious and consider, at least in principle, the full Quantum Gravity path integral, where both the metric $g_{\mu\nu}$ and the matter fields φ fluctuate. The spacetime geometry is therefore dynamical, it can be anything as long as it is non-singular, we should even be ready to sum over different topologies. Is there something we can keep fixed in this context? Yes, the boundary conditions at infinity. This approach has been pioneered in [10] (see e.g. [11] for more details).

We introduce a path integral of the form

$$Z = \int \mathcal{D}g_{\mu\nu} \mathcal{D}\varphi e^{-I_E[g_{\mu\nu}, \varphi]}, \quad (5.40)$$

with some measure $\mathcal{D}g_{\mu\nu}$ for the metric and $\mathcal{D}\varphi$ for the matter fields. Note that this is already in Euclidean signature. There are at least three good reasons for choosing to work in Euclidean rather than Lorentzian signature:

- 1) in general the path integral has better convergence properties;
- 2) we saw that black hole geometries become perfectly regular in Euclidean signature: the space ends at the value of the radial coordinate that in Lorentzian signature corresponds to the position of the event horizon; thus the curvature singularity is excluded from the space. So *going to Euclidean signature allows one to include the contribution of black holes to the path integral while avoiding the curvature singularities that characterize the Lorentzian solutions*;
- 3) we can compute thermal partition functions, which are relevant for black hole physics.

We require that as $r \rightarrow \infty$, locally the space looks like Euclidean flat space. In addition we ask that both the metric and the matter fields are periodic in Euclidean time, with a given period β .

The fact that formally we have written down the path integral does not mean that we are able to compute it. We will see later how this in principle can be done in special situations related to string theory.

One thing we can do is a saddle point approximation around the extrema of the action, namely around the solutions to the classical equations of motion. Adopting the background field method, we split the fields in a background term, solving the classical equations of motion, and a fluctuation term:

$$g = \bar{g} + \delta g, \quad \varphi = \bar{\varphi} + \delta \varphi, \quad (5.41)$$

and expand the classical action as

$$I[g, \varphi] = I[\bar{g}, \bar{\varphi}] + I_2[\delta g, \delta \varphi] + \dots \quad (5.42)$$

where $I[\bar{g}, \bar{\varphi}]$ is the classical on-shell action, while I_2 is quadratic in the fluctuations. The partition function reads

$$-\log Z = I[\bar{g}, \bar{\varphi}] - \log \int \mathcal{D}\delta g \mathcal{D}\delta \varphi e^{-I_2[\delta g, \delta \varphi]} + \dots \quad (5.43)$$

The former is the dominant contribution to the path integral from the saddle point, while the second is a path integral for an action quadratic in the fluctuations, that corresponds to one-loop quantum corrections and is computed by evaluating a functional determinant.¹²

5.5 The Euclidean on-shell action

Let us evaluate the semiclassical contribution of the Schwarzschild black hole to the Euclidean Quantum Gravity path integral.

This is less trivial than what one may think. Since we need to integrate the scalar curvature R , which vanishes for Schwarzschild, we may expect that the result is zero, but in fact there is a crucial contribution from a boundary term to take into account. In order to regulate the long distance divergence that will appear due to the infinite volume of the spacetime, we first assume that the spacetime just extends up to some large but finite value of r , that we call r_0 . This plays the role of a ‘‘cuf-off’’, that can be sent to infinity at the end of the computation. So our spacetime M has a boundary at $r = r_0$, that we denote by ∂M .

The complete Euclidean action on a space with a boundary is

$$I = -\frac{1}{16\pi} \int_M d^4x \sqrt{g} R - \frac{1}{8\pi} \int_{\partial M} \sqrt{h} K, \quad (5.44)$$

¹²In order to see that the classical term is dominant, one should reinstate the factors of \hbar .

where in addition to the familiar Einstein-Hilbert terms there is a boundary term, known as the Gibbons-Hawking-York (GHY) term. Here, h_{ij} is the induced metric on the boundary, and $K = h^{ij}K_{ij}$ is the trace of the extrinsic curvature K_{ij} , defined as

$$K_{ij} = \frac{1}{2} \mathcal{L}_n h_{ij}, \quad (5.45)$$

where \mathcal{L} is the Lie derivative and n is the outward pointing unit vector normal to ∂M . For a metric of the form $ds^2 = N^2 dr^2 + h_{ij} dx^i dx^j$ (we only consider metrics of this form), the extrinsic curvature of the hypersurface $r = r_0$ is simply given by $K_{ij} = \frac{1}{2N} \frac{\partial}{\partial r} h_{ij} \Big|_{r=r_0}$.

The GHY term is needed in order to have a well-definite variational problem with Dirichlet boundary conditions for the metric. The variation of the Einstein-Hilbert term is schematically of the form

$$\delta \int_M d^4x \sqrt{g} R = \int_M (\text{eom}) \delta g + \int_{\partial M} [X(g, \partial g) \delta g + Y(g, \partial g) \partial \delta g], \quad (5.46)$$

where the boundary terms arise from integration by parts. Imposing Dirichlet boundary conditions means that the metric is held fixed at the boundary, namely $\delta g|_{\partial M} = 0$. This makes the first boundary term vanish; however the second term does not vanish in general, so the action would not be extremized upon imposing the equations of motion in the bulk. The Gibbons-Hawking-York term cures this problem: its variation precisely cancels the second boundary term in (5.46), thus leaving us with a good Dirichlet variational problem.

Let us evaluate the action (5.56) for the Euclidean Schwarzschild solution (5.24). Since $R = 0$, the Einstein-Hilbert term vanishes, and the whole contribution is from the boundary term. The induced metric on a hypersurface of constant r is given by

$$h_{ij} dx^i dx^j = \left(1 - \frac{r_s}{r}\right) d\tau^2 + r^2 d\Omega_2, \quad (5.47)$$

and describes the space $S^1 \times S^2$. The trace of the extrinsic curvature, evaluated at $r = r_0$, is

$$K = \frac{2}{r_0} - \frac{r_s}{2r_0^2} + \mathcal{O}(r_0^{-4}), \quad (5.48)$$

and the GHY term evaluates to

$$-\frac{1}{8\pi} \int_{\partial M} \sqrt{h} K = \beta \left(-r_0 + \frac{3}{4} r_s \right) + \dots, \quad (5.49)$$

where the dots denote terms that go to zero when we send $r_0 \rightarrow \infty$. This diverges as we send $r_0 \rightarrow \infty$. So we need to find a good counterterm that subtracts the divergence before sending the cutoff to infinity. The idea is to subtract “the contribution of flat space”, so

that the action of flat space is zero by construction. More precisely, one subtracts the GHY term computed for a boundary surface of identical intrinsic geometry as ∂M , but embedded in flat space. In our case, the appropriate choice for the flat space metric is

$$ds_{\text{flat}}^2 = dr^2 + h_{ij}^{\text{flat}} dx^i dx^j = dr^2 + \left(1 - \frac{r_s}{r_0}\right) d\tau^2 + r^2 d\Omega_2, \quad (5.50)$$

where it is important to notice that $h_{\tau\tau}$ is a fixed constant (in particular, independent of r), so we are just describing $\mathbb{R}^4 = \mathbb{R}_\tau \times \mathbb{R}^3$. Clearly, the metric induced on the hypersurface at $r = r_0$ is identical to the one on ∂M in Schwarzschild. The counterterm evaluates to

$$\frac{1}{8\pi} \int_{\partial M} \sqrt{h^{\text{flat}}} K^{\text{flat}} = \beta \left(r_0 - \frac{r_s}{2}\right) + \dots \quad (5.51)$$

Adding this to (5.49), we see that not only the divergence is removed, but the finite term is also modified. The final result for the renormalized on-shell action reads

$$I_{\text{ren}} = \frac{1}{4} \beta r_s = \pi r_s^2, \quad (5.52)$$

where in the second step we used that the periodicity of the Euclidean time coordinate in the Schwarzschild solution is fixed to $\beta = T_H^{-1} = 4\pi r_s$.

This is the leading contribution to the canonical partition function,¹³

$$-\log Z(\beta) = I_{\text{ren}} = \frac{1}{16\pi} \beta^2. \quad (5.53)$$

Using standard thermodynamics, we deduce the energy

$$E = -\partial_\beta \log Z = \frac{\beta}{8\pi} = M. \quad (5.54)$$

Then the log of the microcanonical partition function, namely the entropy, is obtained as a Legendre transform

$$\begin{aligned} S &= \log Z(\beta) + \beta E \\ &= \frac{\beta^2}{16\pi} = \pi r_s^2 = \frac{A}{4}. \end{aligned} \quad (5.55)$$

We have thus re-derived the Bekenstein-Hawking formula for the black hole entropy by a completely different method.

¹³We can also write $I_{\text{ren}} = -\log Z(\beta) = \beta F$, where F is the free energy.

5.6 The on-shell action in the grand-canonical ensemble

One can also compute the Euclidean on-shell action for the Kerr-Newman black hole. The full Euclidean action, including the counterterm, now is

$$I = -\frac{1}{16\pi} \int_M d^4x \sqrt{g} (R - F_{\mu\nu} F^{\mu\nu}) - \frac{1}{8\pi} \int_{\partial M} \sqrt{h} K + \frac{1}{8\pi} \int_{\partial M} \sqrt{h^{\text{flat}}} K^{\text{flat}}, \quad (5.56)$$

Since the energy-momentum tensor of the Maxwell field is traceless in four dimensions, we still have $R = 0$. For the Maxwell term, we can use

$$\int_M d^4x \sqrt{g} F_{\mu\nu} F^{\mu\nu} = 2 \int_M F \wedge *F = 2 \int_M [d(A \wedge *F) - A \wedge d *F] = 2 \int_{\partial M} A \wedge *F, \quad (5.57)$$

where in the last step we used the Maxwell equation and the Stokes theorem. So again the action reduces to a boundary term. Evaluating this boundary term carefully in the gauge where the gauge potential is regular, one finds

$$I = \frac{\beta}{2} (M - \Phi Q). \quad (5.58)$$

As we already discussed, we should consider ourselves in the grand-canonical ensemble, where the inverse temperature β , the angular potential Ω and the electric potential Φ can be obtained by analyzing the Euclidean section of the solution. Therefore the on-shell action should provide minus the logarithm of the grand-canonical partition function,

$$\log Z(\beta, \Omega, \Phi) = -I. \quad (5.59)$$

Recalling the generalized Smarr relation (2.57), we can write

$$\frac{\text{Area}}{4} = \beta \left(\frac{1}{2} M - \Omega J - \frac{1}{2} \Phi Q \right) = -I + \beta (M - \Omega J - \Phi Q). \quad (5.60)$$

We have thus obtained

$$\frac{\text{Area}}{4} = \log Z + \beta (M - \Omega J - \Phi Q). \quad (5.61)$$

One also verifies that¹⁴

$$J = \frac{1}{\beta} \frac{\partial \log Z}{\partial \Omega} \Big|_{\beta, \Phi}, \quad Q = \frac{1}{\beta} \frac{\partial \log Z}{\partial \Phi} \Big|_{\beta, \Omega}, \quad M = - \frac{\partial \log Z}{\partial \beta} \Big|_{\Omega, \Phi} + \Omega J + \Phi Q. \quad (5.62)$$

¹⁴Checking these relations is not immediate because we do not have the expressions for the charges $\{M, J, Q\}$ as functions of the potentials $\{\beta, \Omega, \Phi\}$ at hand. On the other hand, it is easy to express the potentials as functions of the charges. Denoting by $p^i = \{\beta, \Omega, \Phi\}$ the vector of potentials and by $c^j = \{M, J, Q\}$ the vector of charges, the relations (5.62) are most easily checked by first computing the Jacobian $J^i_j = \frac{\partial p^i(c)}{\partial c^j}$ and then evaluating its inverse to obtain the derivatives $\frac{\partial}{\partial p^i} = (J^{-1T})^j_i \frac{\partial}{\partial c^j}$.

These relations tell us that $\frac{\text{Area}}{4}$ is the Legendre transform of the logarithm of the grand-canonical partition function $Z(\beta, \Omega, \Phi)$ with respect to its variables. This is precisely the definition of the logarithm of the microcanonical partition function, namely the entropy.

The Euclidean approach thus shows that the $T = \frac{\kappa}{2\pi}$ and $S = \frac{\text{Area}}{4}$ laws also hold for the Kerr-Newman solution. These are in fact very universal relations.

6 Black holes in AdS and phase transitions

What happens if there is more than one solution to the classical equations of motion satisfying the same prescribed boundary conditions? Each solution will provide a saddle of the gravitational partition function and will thus contribute to it. For instance, for the case where there are two such solutions, sol1 and sol2, the partition function in the semiclassical approximation reads

$$Z \simeq e^{-I[\text{sol1}]} + e^{-I[\text{sol2}]} . \quad (6.1)$$

The solution with least action will dominate the statistical ensemble. Indeed, suppose $I[\text{sol1}] < I[\text{sol2}]$; then

$$Z \simeq e^{-I[\text{sol1}]} (1 + e^{I[\text{sol1}] - I[\text{sol2}]}) . \quad (6.2)$$

is approximated by $e^{-I[\text{sol1}]}$, up to an contribution that is exponentially suppressed in the semiclassical approximation where $\hbar \rightarrow 0$.

It can happen that different solutions dominate in different regimes of the variables characterizing the statistical ensemble considered (in the grand-canonical ensemble, these are e.g. the temperature, the angular potential, the electric potential). In this case there must be a *phase transition* between the different regimes.

An emblematic example is the Hawking-Page phase transition for black holes in AdS [12], which also has a beautiful interpretation in the context of the AdS/CFT correspondence [13]. The papers [12, 13] are very clearly written, so we directly refer to them and do not report their content in these notes.

7 Wald's entropy

So far we only considered two-derivative theories, such as GR coupled to a Maxwell field, possibly with a cosmological constant. However we know that GR should be seen as an effective field theory, and as such in the spirit of effective field theories it has to be corrected

by higher derivative terms suppressed by the Planck scale, schematically

$$S = M_P^2 \int d^4x \sqrt{-g} \left(R + \frac{1}{M_P^2} R_{\mu\nu\rho\sigma}^2 + \frac{1}{M_P^4} R_{\mu\nu\rho\sigma}^4 + \dots \right). \quad (7.1)$$

While the two-derivative Einstein-Hilbert term is universal, the precise form of the higher-derivative terms depends on the UV completion of the theory. In particular, string theory determines an infinite series of higher-derivative terms, only some of which are known.

In the presence of higher-derivative terms, the second law of black hole mechanics is in general not satisfied, so it may be that the interpretation of black holes as thermodynamic objects is only valid in the limiting low-energy situation where only the two-derivative action matters. However *Wald showed that one can still associate an entropy to black holes in higher derivative theories of gravity, that satisfies the first law* [14, 15, 16].

In Wald's formulation, the black hole entropy is related to the Noether charge of diffeomorphisms under the Killing vector field which generates the event horizon of a stationary black hole. Given a generally covariant action I including higher-derivative terms, Wald's formula for the entropy S reads

$$S = 2\pi \int_{\mathcal{S}} \text{vol}_{\mathcal{S}} \frac{\delta I}{\delta R_{\mu\nu\rho\sigma}} \epsilon_{\mu\rho} \epsilon_{\nu\sigma}, \quad (7.2)$$

where $\epsilon^{\mu\nu}$ is binormal to the horizon and $\text{vol}_{\mathcal{S}}$ is the volume form induced on the intersection \mathcal{S} of the horizon with a spacelike hypersurface. The variation of the action with respect to the Riemann tensor $R_{\mu\nu\rho\sigma}$ must be performed by first expressing all possible antisymmetrizations of covariant derivatives appearing in the action in terms of the Riemann tensor (so that only symmetric combinations of covariant derivatives remain), and then treating the Riemann tensor as an independent variable.

We will not directly use this formula, but rather rely on a simpler approach valid for extremal black holes.

8 The quantum entropy of extremal black holes

Sen developed a method for computing the Wald entropy of extremal black holes, which conveniently exploits the enhanced symmetry of their near-horizon field configuration. This is still in a classical effective theory of gravity, though with higher derivatives. Then he went further and proposed a concrete (and computable) definition for the entropy in the full Quantum Gravity theory. Two of Sen's original papers are [17, 18]; nice reviews can be found in [19, 20, 21].

8.1 Extremal black holes

Recall that when we discussed the Kerr-Newman solution, we assumed

$$M^2 \geq a^2 + P^2 + Q^2, \quad M > 0, \quad (8.1)$$

so that the roots

$$r_{\pm} = M \pm \sqrt{M^2 - (a^2 + P^2 + Q^2)} \quad (8.2)$$

of the polynomial $\Delta(r)$ are real and positive. Both the black hole temperature and entropy depend on r_{\pm} , so it is crucial that these are well defined. When the bound (8.1) is saturated, namely when

$$M = \sqrt{a^2 + P^2 + Q^2}, \quad (8.3)$$

we say that we have an *extremal black hole*. This corresponds to asking that the inner and outer horizons coincide,

$$r_+ = r_- = r_*, \quad \text{with} \quad r_* = M = \sqrt{a^2 + P^2 + Q^2}. \quad (8.4)$$

Because $r_+ - r_- = 0$, the surface gravity vanishes and the black hole is at zero temperature. This means that it does not radiate. However the area of the horizon

$$\frac{A}{4} = \pi (r_*^2 + a^2) \quad (8.5)$$

does not vanish, hence the black hole still carries a non-zero entropy.

The fact that extremal black holes are stable against evaporation but still carry a large entropy allows us to separate the problem of studying the microscopic origin of the black hole entropy from the one of understanding Hawking radiation. Extremal black holes are isolated quantum systems, while radiating black holes are in equilibrium with a thermal bath, so they are not really isolated. Moreover, since the temperature is zero, the entropy should just count the degeneracy of *ground* states (with assigned charges J, P, Q). For the rest of these lectures we will focus on the problem of accounting for the entropy of extremal black holes.

For simplicity, we take $a = 0$ in the Kerr-Newman solution, namely we focus on the dyonic Reissner-Nordström solution to the Einstein-Maxwell theory (2.59). The solution reads

$$ds^2 = - \left(1 - \frac{r_-}{r}\right) \left(1 - \frac{r_+}{r}\right) dt^2 + \frac{dr^2}{\left(1 - \frac{r_-}{r}\right) \left(1 - \frac{r_+}{r}\right)} + r^2 d\Omega_2, \quad (8.6)$$

$$F = \frac{Q}{r^2} dr \wedge dt - P \sin \theta d\theta \wedge d\phi. \quad (8.7)$$

If first we impose the extremality condition $r_{\pm} = r_* = M = \sqrt{Q^2 + P^2}$ and then take a near horizon limit setting $r = r_*(1 + \rho)$, we obtain at leading order as $\rho \rightarrow 0$:

$$ds^2 = -\rho^2 dt^2 + r_*^2 \frac{d\rho^2}{\rho^2} + r_*^2 d\Omega_2 + \dots, \quad (8.8)$$

The Rindler factor that we obtained in the near-horizon limit of Schwarzschild is replaced here by AdS₂. This means that we don't have to impose periodicity of the Euclidean time, because AdS₂ does not cap off at finite distance, it rather has an infinite throat. This can be seen by making the change of coordinate $\rho = e^\sigma$; the range of σ is the whole real line, and the space never ends.¹⁵ Since the Euclidean time is not periodically identified, there is no finite temperature. However, we can define a thermodynamics for extremal black holes starting from the finite temperature case and taking the limit. It is in this limiting sense that the thermodynamics of extremal black holes should be understood.

It is convenient to define a slightly different scaling limit of the Reissner-Nordström solution that zooms in on the near-horizon region and at the same time leads to extremality. Transform t, r into new (dimensionless) coordinates \tilde{t}, \tilde{r}

$$t = r_+^2 \frac{\tilde{t}}{\lambda}, \quad r = r_+ + \lambda(\tilde{r} - 1), \quad (8.9)$$

where the (dimensionful) parameter λ measures the distance between the inner and outer horizons,

$$r_- = r_+ - 2\lambda, \quad (8.10)$$

or in other words it tells us how far we are from extremality. Note that the positions of the inner and outer horizons $r = r_{\pm}$ corresponds to $\tilde{r} = \pm 1$ in the new coordinate. The Reissner-Nordström solution becomes

$$ds^2 = -\frac{r_+^4(\tilde{r}^2 - 1)}{(r_+ + \lambda(\tilde{r} - 1))^2} d\tilde{t}^2 + (r_+ + \lambda(\tilde{r} - 1))^2 \left(\frac{d\tilde{r}^2}{\tilde{r}^2 - 1} + d\Omega_2 \right),$$

$$F = \frac{Q r_+^2}{(r_+ + \lambda(\tilde{r} - 1))^2} d\tilde{r} \wedge d\tilde{t} - P \sin \theta d\theta \wedge d\phi. \quad (8.11)$$

We can now take the extremal limit by sending $\lambda \rightarrow 0$, which implies $r_{\pm} \rightarrow r_* = \sqrt{Q^2 + P^2}$. In this way we obtain

$$ds^2 = r_*^2 \left[-(\tilde{r}^2 - 1) d\tilde{t}^2 + \frac{d\tilde{r}^2}{\tilde{r}^2 - 1} + d\Omega_2 \right],$$

$$F = Q d\tilde{r} \wedge d\tilde{t} - P \sin \theta d\theta \wedge d\phi. \quad (8.12)$$

¹⁵In the original coordinate, this is seen by checking that the proper length of a line of constant θ, ϕ, t extending from $r = r_0$ to $r = r_*$ is $\int_{r_*}^{r_0} \frac{dr}{1 - r_*/r} = \infty$.

Since (8.11) is a solution to the equations of motion for any value of λ , the limiting configuration (8.12) is also a solution. This scaling limit also has the virtue of keeping the two horizons at finite distance, so that the solution still looks like a black hole after taking the limit. This will be important in the following, in particular when we will discuss the regularity conditions of the Euclidean section of the solution.

All known extremal black hole solutions have an AdS_2 factor in the near-horizon geometry. It can also be proven that the converse is true under mild assumptions [22]. The rest of the near-horizon geometry is a compact manifold \mathcal{M}_{d-2} that in general may be fibered over AdS_2 . The $SO(2, 1) \simeq SL(2)$ isometry of AdS_2 is a symmetry of the near-horizon solution, in the sense that all fields are invariant under it. By contrast, $SO(2, 1)$ is not a symmetry of the original solution: it only arises in the near-horizon geometry as an enhancement of time translation invariance.

We will take the presence of an *AdS_2 factor in the near-horizon geometry as a definition of extremal black holes*, in any generally covariant theory of gravity, including all sort of higher derivative terms.

8.2 The entropy function

Exploiting wisely the symmetries of the extremal near-horizon geometry, Sen obtained a simplified way to express the Wald entropy, that also paved the way for defining the full quantum entropy.

Consider an arbitrary theory of gravity in four spacetime dimensions (this can be generalized to other dimensions) coupled to $U(1)$ gauge fields $A_\mu^{(i)}$, $i = 1, \dots, \text{rank } G$, and neutral scalar fields ϕ_s , with $s = 1, \dots, N$. There could also be fermion fields, that will play no role in our discussion as they are always set to zero in the solution. This theory may contain higher derivative terms and come from compactification of string theory, for instance. The action reads

$$I = \int d^4x \sqrt{-g} \mathcal{L}, \quad (8.13)$$

where \mathcal{L} is a general coordinate invariant and local Lagrangian. We could also think of dimensionally reducing the four-dimensional theory on the compact manifold \mathcal{M}_2 to a two-dimensional gravity theory. A priori the dimensional reduction is not a truncation, i.e. we should keep the infinite set of modes of the higher-dimensional fields on the internal space. From this point of view, the action reads

$$I = \int dt dr \sqrt{-g^{(2)}} \mathcal{L}^{(2)}, \quad \text{with} \quad \mathcal{L}^{(2)} = \int_{\mathcal{M}_2} \text{vol}_{\mathcal{M}_2} \mathcal{L} \quad (8.14)$$

and $g^{(2)}$ is the determinant of the 2d metric.

For simplicity we will discuss a static solution, where $\mathcal{M}_2 = S^2$, endowed with the round metric (many generalizations are possible, including rotating black holes, asymptotically AdS black holes, different horizon topologies, etc.). A static extremal black hole will have a near-horizon geometry $\text{AdS}_2 \times S^2$, with $\text{SO}(2,1) \times \text{SO}(3)$ symmetry. This means that the fields must take the form¹⁶

$$\begin{aligned} ds^2 &= v_1 \left(-(r^2 - 1) dt^2 + \frac{dr^2}{r^2 - 1} \right) + v_2 d\Omega_2, \\ F^{(i)} &= e_i dr \wedge dt + p_i \sin \theta d\theta \wedge d\phi, \\ \phi_s &= u_s, \end{aligned} \tag{8.15}$$

where $F^{(i)} = dA^{(i)}$. The only variables here are the constants v_1, v_2, e_i, p_i, u_s , all the rest being fixed by symmetries. The e_i and p_i parameterize the near-horizon electric and magnetic fields, respectively.

From the point of view of the dimensional reduction to 2d, we are keeping just the constant modes of the fields on S^2 , the extremal near-horizon configuration is just an AdS_2 vacuum solution of the 2d theory with radius controlled by v_1 , while the e_i parameterize the 2d gauge field strengths, v_2, u_s are the constant values of 2d scalar fields, and the p_i are coupling constants coming from “flux parameters” in the internal S^2 geometry. The constants v_1, v_2, u_s need to be determined using the equations of motion, which in this background reduce to a set of algebraic equations.

Plugging (8.15) into the Lagrangian and integrating over the angular coordinates, the 2d Lagrangian becomes

$$\mathcal{L}^{(2)}|_{\text{AdS}_2} = v_2 \int_{S^2} d\theta d\phi \sin \theta \mathcal{L}|_{\text{AdS}_2 \times S^2} = 4\pi v_2 \mathcal{L}|_{\text{AdS}_2 \times S^2} \tag{8.16}$$

and the 2d Lagrangian density evaluates to

$$f = \sqrt{-g^{(2)}} \mathcal{L}^{(2)}|_{\text{AdS}_2} = v_1 \mathcal{L}^{(2)}|_{\text{AdS}_2} = 4\pi v_1 v_2 \mathcal{L}|_{\text{AdS}_2 \times S^2}, \tag{8.17}$$

This is independent of t, r , while it depends on the various constants,

$$f = f(u, v, e, p). \tag{8.18}$$

It just remains to introduce the *entropy function*

$$\mathcal{E}(u, v, e; p, q) = 2\pi (e_i q_i - f(u, v, e, p)). \tag{8.19}$$

¹⁶Here we drop the tildes on the radial and time coordinate introduced in Eq. (8.12).

In this function, u, v, e are the variables, while p, q should be seen as fixed parameters.

Now the claim is that the Wald entropy is computed by extremizing the entropy function with respect to the variables e, u, v , and evaluating it at the extremum. The extremization equations

$$\frac{\partial \mathcal{E}}{\partial v_1} = \frac{\partial \mathcal{E}}{\partial v_2} = 0, \quad \frac{\partial \mathcal{E}}{\partial u_s} = 0, \quad (8.20)$$

are equivalent to imposing the Einstein equation and the equations for the scalar fields ϕ_s , respectively. The Maxwell equations are trivially satisfied by the ansatz, however we can extract some more information regarding the gauge fields. Extremizing with respect to e_i gives

$$\frac{\partial \mathcal{E}}{\partial e_i} = 0 \quad \Leftrightarrow \quad q_i = \frac{\partial f}{\partial e_i}. \quad (8.21)$$

This is just telling us that the new parameters q_i introduced in (8.19) are identified with the electric charges of the black hole: indeed in general the electric charge is defined as $q = \int_{S^2} \frac{\delta I}{\delta F_{rt}}$. This is also the electric charge of the full black hole solution, as the integral defining the electric charge can be evaluated near the horizon or at infinity, giving the same result.

The extremization equations above generically determine the near-horizon values of the e, v, u variables in terms of the electric and magnetic charges q, p . Once these equations are solved, the near-horizon solution is determined. The extremum value

$$\mathcal{E}_*(p, q) = \mathcal{E}(u_*(p, q), v_*(p, q), e_*(p, q); p, q) \quad (8.22)$$

is just a function of the electric and magnetic charges. One can show that the Wald entropy is precisely this extremum value,

$$S_{\text{Wald}}(p, q) = \mathcal{E}_*(p, q). \quad (8.23)$$

The proof requires some work, and can be found in [19].

The example of Reissner-Nordström. Let us illustrate the entropy function formalism described above by computing the entropy in the simple case of the Reissner-Nordström black hole solution to the Maxwell-Einstein theory. In this case, we find

$$\sqrt{-g} = v_1 v_2 \sin \theta, \quad R = -\frac{2}{v_1} + \frac{2}{v_2}, \quad F_{\mu\nu} F^{\mu\nu} = -\frac{2e^2}{v_1^2} + \frac{2p^2}{v_2^2}, \quad (8.24)$$

so

$$\begin{aligned} f &= \frac{1}{16\pi} \int_{S^2} d\theta d\phi \sqrt{-g} (R - F_{\mu\nu} F^{\mu\nu}) \\ &= \frac{1}{2} \left(v_1 - v_2 + \frac{v_2}{v_1} e^2 - \frac{v_1}{v_2} p^2 \right) \end{aligned} \quad (8.25)$$

and the entropy function $\mathcal{E} = 2\pi(eq - f)$ is extremized at

$$v_1 = v_2 = q^2 + p^2, \quad e = q, \quad (8.26)$$

which is in agreement with the near-horizon solution (8.12) (upon identifying $q = Q, p = -P$). The value of \mathcal{E} at the extremum is

$$\mathcal{E}_* = \pi(q^2 + p^2) = S_{\text{BH}}, \quad (8.27)$$

in agreement with the entropy of the extremal Reissner-Norsdström black hole.

Attractor mechanism. Some of the scalar fields in the theory may be flat directions of the scalar potential in the Lagrangian. In this case they take arbitrary values in the Minkowski vacuum and are called *moduli*. In an asymptotically flat black hole solution, these scalar fields take arbitrary values at infinity. Since these asymptotic values do not enter in the definition of the entropy function, they will not affect its extremization problem either. It follows that the near-horizon values of the fields that enter in the entropy function, as well as the black hole entropy, are completely independent of the moduli. This result generalizes the *attractor mechanism* first observed in $\mathcal{N} = 2$ supergravity [23, 24, 25].

8.3 Relation with Euclidean on-shell action

We now show that the entropy function \mathcal{E} is closely related to the Euclidean on-shell action of the gravitational theory. In turn, the Euclidean on-shell action can be seen as the saddle point value of the gravitational partition function Z in the semiclassical approximation,

$$Z \simeq e^{-I_E}, \quad (8.28)$$

where the partition function is defined with prescribed AdS_2 boundary conditions.

In order to demonstrate the relation with the Euclidean action, we take a two-dimensional point of view and aim at computing

$$I_E = I_{\text{bulk}} + I_{\text{dry}}. \quad (8.29)$$

Here,

$$I_{\text{bulk}} = - \int d\tau dr \sqrt{g^{(2)}} \mathcal{L}^{(2)}|_{\text{AdS}_2} = - \int d\tau dr f, \quad (8.30)$$

where $\tau = it$ is the Euclidean time and the 2d Lagrangian density evaluated on the near-horizon solution is just the function f introduced in (8.17). I_{bdry} denotes boundary terms that will be needed to remove the divergences of the bulk action.

To compute the integral we first need to identify the appropriate range of the coordinates. The Euclidean 2d field configuration is

$$\begin{aligned} ds^2 &= v \left((r^2 - 1) d\tau^2 + \frac{dr^2}{r^2 - 1} \right), \\ F^{(i)} &= -i e_i dr \wedge d\tau \quad \Leftrightarrow \quad A^{(i)} = -i e_i (r - 1) d\tau, \\ \phi_s &= u_s, \end{aligned} \tag{8.31}$$

where the gauge for $A^{(i)}$ has been fixed by regularity: it is chosen in such a way that $A^{(i)} \rightarrow 0$ as $r \rightarrow 1$, since the differential $d\tau$ is not well defined there. Inspection of the metric shows that the Euclidean time needs to be identified as $\tau \sim \tau + 2\pi$. This is easily seen by changing the radial coordinate as $r = \cosh \eta$, so that the metric becomes

$$ds^2 = v (\sinh^2 \eta d\tau^2 + d\eta^2), \tag{8.32}$$

and requiring regularity for $\eta \rightarrow 0$.

The range of the radial coordinate is a priori $1 \leq r \leq \infty$. This of course leads to an infinite volume of AdS_2 , so our action I is a priori divergent. We can regularize the volume assuming that the integral over r is performed only up to $r = r_0 < \infty$, and then subtract the long-distance divergence.

Our bulk action thus evaluates to

$$I_{\text{bulk}} = - \int_0^{2\pi} d\tau \int_1^{r_0} dr f = -2\pi(r_0 - 1)f. \tag{8.33}$$

The divergence as $r_0 \rightarrow \infty$ can be removed by an appropriate choice of boundary terms I_{bdry} . One can show (see [20]) that any boundary term that is local in the boundary fields and gauge invariant can only contribute with a $\mathcal{O}(r_0)$ term and with terms that vanish when $r_0 \rightarrow \infty$, but not with a $\mathcal{O}(1)$ term. The $\mathcal{O}(r_0)$ contribution can be chosen so as to cancel the divergence in the bulk action. On the other hand, the finite term in the bulk action is not affected by local, gauge invariant boundary terms and is thus non-ambiguous. We will take such finite term as the definition of the renormalized bulk action. We conclude that

$$I_E(e, p) = 2\pi f_*(e, p), \tag{8.34}$$

where the $*$ indicates that we have extremized f with respect to the variables u_s, v . Then the partition function in the semiclassical approximation evaluates to

$$Z(e, p) \simeq e^{-I_E(e, p)} = e^{-2\pi f_*(e, p)}. \tag{8.35}$$

In order to interpret this partition function we should discuss the boundary conditions for the gauge field. We notice that the action is extremized upon imposing the equations of motion only if the gauge field A is held fixed at the boundary in the variational problem. For instance, the variation of a term of the type $\int_M \mathcal{G}_{ij} F^{(i)} \wedge *F^{(j)}$ in the action yields the boundary term $\int_{\partial M} \delta A^{(i)} \wedge \mathcal{G}_{ij} *F^{(j)}$, which vanishes if we set $\delta A^{(i)} = 0$ on ∂M . This boundary condition is natural in asymptotically AdS $_d$ spacetimes of dimension $d > 3$. The reason is that for $d > 3$ the asymptotic solution to the Maxwell equation in the radial gauge $A_r = 0$ is of the form $A = a_0 + \frac{a_1}{r^{d-3}} + \dots$, hence the asymptotic value of the gauge field $A \rightarrow a_0$ is the dominating (non-normalizable) mode, while the asymptotic value of the field strength component $F_{r\mu}$ is controlled by the subleading (normalizable) mode a_1 . Since it is the dominating term, it is natural that the boundary gauge field a_0 is kept fixed in the variational problem, rather than the field strength. Since a_0 contains the electric potential Φ in its temporal component, the boundary condition just described leads us to interpret the Euclidean on-shell action as (minus the logarithm of) a grand-canonical partition function, describing a statistical ensemble where the electric potential Φ is held fixed, while the electric charge is determined dynamically by regularity of the bulk solution. This is also the interpretation of the 2d partition function obtained in (8.35). Note indeed that for the gauge fields (8.31), the constant mode is proportional to e_i .

In 2d, however, the boundary condition where the r -independent mode is held fixed in the variational problem is not natural, and generically causes problems in the path integral. The reason is that the solution of the Maxwell equation in an asymptotically AdS $_2$ space has an asymptotic behavior of the type

$$A_t = e r + \Phi + \mathcal{O}(r^{-1}), \quad (8.36)$$

where again we are assuming the gauge $A_r = 0$. So in 2d the term controlling the field strength is dominating over the Φ term. It is thus more natural to keep the field strength, that is the electric charge, fixed, and allow the potential Φ to fluctuate. In other words, in 2d it is more natural to set ourselves in the microcanonical ensemble, rather than in the grand-canonical ensemble. This is achieved by adding the following boundary term to the Euclidean action

$$I_E \rightarrow \widehat{I}_E = I_E + i q_i \int d\tau A_\tau^{(i)}, \quad (8.37)$$

which precisely cancels the boundary term containing $\delta A^{(i)}$ in the variation of the bulk action. The new boundary term should also be understood with the prescription that only the finite part is kept as $r_0 \rightarrow \infty$. For the gauge field (8.31) we have

$$i q_i \int d\tau A_\tau^{(i)} = 2\pi q_i e_i (r_0 - 1). \quad (8.38)$$

Removing the divergent term by a choice of boundary terms, we arrive at the finite result

$$\widehat{I}_E(q, p) = -2\pi(e_i q^i - f)_* \equiv -\mathcal{E}_*(q, p), \quad (8.39)$$

where the $*$ indicates that we have performed extremization with respect to the variables u_s, v , as well as e_i . The corresponding gravitational partition function then reads in the semiclassical approximation

$$\log \widehat{Z}(q, p) = -\widehat{I}_E(q, p) = \mathcal{E}_*(q, p) = S_{\text{Wald}}(q, p). \quad (8.40)$$

We have thus recovered the entropy function \mathcal{E}_* , and therefore Wald's formula for the black hole entropy. This demonstrates that the entropy function really is a renormalized Euclidean on-shell action. The $q_i e_i$ term comes from a Wilson line implementing the appropriate boundary condition for the gauge field that keeps the electric charge fixed and allows the constant mode of the gauge field to fluctuate.

The same result is obtained by taking a Legendre transform of the logarithm of the grand-canonical partition function $\log Z(e, p)$, which replaces the variables e_i with the electric charges q_i . Recall that the e_i also control the constant mode of the gauge field in (8.31), namely the electric potentials Φ_i that are the thermodynamic variables conjugate to the electric charges q_i . However, for the purpose of promoting this saddle point evaluation to a full path integral in the next section, it is more natural to work with the microcanonical partition function rather than with the grand-canonical partition function.

To summarize: starting from a 2d gravitational theory including all sort of higher derivative terms, we computed the saddle point value of the microcanonical partition function $\log \widehat{Z}(q, p) = -\widehat{I}_E(q, p)$ by evaluating the renormalized Euclidean action $\widehat{I}_E(q, p)$ of an AdS_2 solution with the boundary condition that the electric charge is held fixed, rather than the potential. We have obtained the chain of equalities (8.40), which expresses the equivalence between $\log \widehat{Z}(q, p)$, the extremum value \mathcal{E}_* of the entropy function, and Wald's formula for the black hole entropy.

8.4 Quantum entropy

So far we have considered higher derivative terms in the effective action, but we have not discussed quantum effects within this effective action. Starting from the classical formulation above, Sen went further and defined a full path integral for the quantum black hole entropy. Sen's quantum black hole entropy is defined as the logarithm of the microcanonical partition

function

$$\begin{aligned}\widehat{Z} &:= \left\langle e^{-iq_i \int d\tau A_\tau^{(i)}} \right\rangle_{\text{AdS}_2}^{\text{finite}} \\ &= \int \mathcal{D}(\text{all fields}) e^{-I_E - iq_i \int d\tau A_\tau^{(i)}} \Big|_{\text{AdS}_2}^{\text{finite}},\end{aligned}\tag{8.41}$$

where “finite” denotes a renormalization of the long-distance divergences analogous to the one discussed in the computation of the classical on-shell action, which only keeps the r_0 -independent term as the radial cutoff r_0 is sent to infinity. The specification “AdS₂” indicates that in the path integral the fields are allowed to fluctuate, but asymptotically they need to have the same behavior that we have seen when discussing the classical case.

It is easy to see that the semiclassical approximation to the gravitational partition function we discussed previously does apply to the definition (8.41) of the quantum entropy. In particular, if we take the large-size limit $v \rightarrow \infty$, the path integral is dominated by the entropy function \mathcal{E}_* .

The path integral (8.41) is in general very hard to compute. However in favourable circumstances interesting results have been obtained. In particular, in the presence of supersymmetry one can

- evaluate more easily logarithmic corrections; these arise as one-loop contributions of massless fields;
- use supersymmetric indices to compute the entropy on the microscopic side;
- exploit the technique of supersymmetric localization directly in supergravity. This gives intriguing results, see e.g. [26, 27, 28].

9 Black hole microstate counting

So far we have discussed the *macroscopic* side of the black hole entropy problem: first we formulated black hole thermodynamics in GR and introduced the Bekenstein-Hawking entropy; then we explored how the black hole entropy can be defined once higher derivative and quantum corrections to GR are taken into account. Contrarily to the universal Bekenstein-Hawking entropy, the form of the higher derivative and quantum corrections depends on the details of the UV complete theory, however we have discussed a formalism allowing to incorporate such corrections on general grounds. This led to Sen’s definition of the quantum entropy for an extremal black hole as a path integral with AdS₂ boundary conditions,

and we argued that in certain circumstances this can (at least in part) be computed in the gravitational theory.

We now briefly discuss the *microscopic* side of the black hole entropy problem. Given the entropy of a thermodynamical system, we can ask what is its microscopic origin, namely how to express the entropy as

$$S = k_B \log d_{\text{micro}}, \quad (9.1)$$

where d is the degeneracy of states in the microcanonical ensemble. These states have assigned value of the energy and of the other charges, such as the angular momentum and the electric charge,

$$S(E, J, Q) = k_B \log d(E, J, Q). \quad (9.2)$$

For instance, one can derive the macroscopic entropy and the other thermodynamic properties of a gas by averaging over a system of many particles, where each microscopic constituent obeys the fundamental laws of mechanics. Similarly, a main challenge for a fundamental theory of quantum gravity is to provide a microscopic derivation of black hole thermodynamics, and in particular to account for the many microstates that make the black hole entropy. We would like to define an appropriate ensemble of states and argue that their degeneracy at fixed value of the charges reproduces the black hole entropy. In order to do this, we will exploit the AdS₂/CFT₁ correspondence.

A first aim is to reproduce the universal Bekenstein-Hawking entropy for extremal (in particular, supersymmetric) black holes as a function of the conserved charges E, J, Q . Secondly, one would like to also reproduce the higher-derivative and quantum corrections to the leading Bekenstein-Hawking term. This corresponds to evaluating a microcanonical partition function. As we have seen, extremal black holes have an AdS₂ factor in the near-horizon geometry. Using the AdS₂/CFT₁ correspondence, the problem of computing the gravitational partition function with AdS₂ boundary conditions, Z_{AdS_2} , can be mapped into the one of computing the partition function Z_{CFT_1} of a conformal quantum mechanics, that is quantum mechanics with an SL(2) global symmetry.¹⁷ The AdS/CFT correspondence indeed states the equivalence between these two partition functions,

$$Z_{\text{AdS}_2} = Z_{\text{CFT}_1}. \quad (9.3)$$

The CFT₁ describes the low-energy limit of the brane system that corresponds to the black hole, and lives at the boundary $r = r_0$; the boundary is just the S^1 parameterized by the Euclidean time τ . It is convenient to rescale this coordinate as $\tau^{\text{new}} = r_0 \tau^{\text{old}}$, so that the

¹⁷This discussion follows [20, Section 4], to which we refer for more details.

period of the new Euclidean time is $\beta = 2\pi r_0$. The partition function in the ensemble where the charges are fixed may be written as the trace of the evolution operator defined by the CFT_1 Hamiltonian,

$$\widehat{Z}_{\text{CFT}_1} = \text{Tr} e^{-\beta H_{\text{CFT}}} . \quad (9.4)$$

Assuming that H_{CFT} is positive definite, for $r_0 \rightarrow \infty$ the partition function is approximated by

$$\widehat{Z}_{\text{CFT}_1} \longrightarrow e^{-2\pi r_0 E_0} d(q) , \quad (9.5)$$

where E_0 is the ground state energy and $d(q)$ denotes the degeneracy of ground states carrying charges q_i . The ground state energy can be set to zero by adding a local counterterm. This parallels the subtraction of divergences that is prescribed in the gravitational path integral (8.41). Comparing (8.41) and (9.5) we arrive at

$$d(q) = \left\langle e^{-iq_i \int d\tau A_\tau^{(i)}} \right\rangle_{\text{AdS}_2}^{\text{finite}} , \quad (9.6)$$

namely the degeneracy of ground states in the CFT_1 reproduces the quantum entropy. We stress again that this result is a direct consequence of the $\text{AdS}_2/\text{CFT}_1$ correspondence.

So the microscopic computation of the quantum entropy consists of enumerating the ground states of the corresponding CFT_1 . How easy is to compute this degeneracy? One issue is that in the regime where the gravitational theory is weakly coupled and it makes sense to speak of the black hole, the CFT_1 is strongly coupled (this is a general feature of the AdS/CFT correspondence), and we know very little of it. However in certain favourable circumstances supersymmetry comes to the rescue, and ensures that the quantity of interest can equally well be computed in the weakly coupled CFT_1 . This is possible when the quantity of interest can be rewritten as a *Witten index*, which is independent of the value of the coupling constants in the theory and is thus protected against quantum corrections. The Witten index is defined as

$$\text{Index} = \text{Tr} (-1)^F e^{-\beta H_{\text{CFT}}} , \quad (9.7)$$

where here we assume that $H_{\text{CFT}} = \{\mathcal{Q}, \overline{\mathcal{Q}}\}$ is a supersymmetric Hamiltonian (\mathcal{Q} is a supercharge in the CFT_1) and F is the fermion number. Since it counts the fermionic states with a minus sign, the index is in principle different from the degeneracy $d(q)$ we are after, where both bosonic and fermionic states contribute to the trace with a positive sign. However, we observe that by the spin-statistics theorem, we can write $(-1)^F = (-1)^{2J}$, where J is the angular momentum for rotation along the azimuthal angle in the S^2 that appears in the black hole near-horizon geometry. Indeed bosonic states have integer angular momentum while

fermionic states have half-integer angular momentum. But here we are considering non-rotating black holes, so the average value of the angular momentum should be zero. Since we are in the microcanonical ensemble, it follows that each microstate must be non-rotating too, namely $J = 0$ for each black hole microstate. We thus expect that the CFT_1 ground states whose degeneracy is $d(q)$ have vanishing angular momentum, and are thus bosonic states. For these states, the index agrees with the degeneracy $d(q)$ and therefore provides a microscopic evaluation of the quantum black hole entropy.

The validity of this approach has been verified in detail in a number of examples. Related ideas can be applied to the evaluation of the entropy of supersymmetric black holes that are asymptotically AdS_{d+1} using the $\text{AdS}_{d+1}/\text{CFT}_d$ correspondence. See [29] for a review of some of the recent developments.

A Statistical ensembles

In this appendix, we review the microcanonical, canonical and grand-canonical ensembles of Statistical Mechanics. We consider a system whose states are characterized by the energy E , the angular momentum J and the electric charge Q . For simplicity we will assume an ensemble with a finite number of states, labelled by the discrete index i . The formulae can easily be adapted to the case where there is a continuum of states, or carrying different quantum numbers.

A.1 Microcanonical ensemble

In the *microcanonical ensemble*, all states i have the same fixed values of E, J, Q , and are assigned equal probability $P_i = 1/Z_{\text{micro}}$, where the microcanonical partition function $Z_{\text{micro}}(E, J, Q)$ is simply the total number of states. Its logarithm is the entropy:

$$S(E, J, Q) = - \sum_i P_i \log P_i = \log Z_{\text{micro}}. \quad (\text{A.1})$$

In this ensemble, the potentials (β, Ω, Φ) conjugate to the charges (E, J, Q) are obtained as

$$\beta = \left. \frac{\partial S}{\partial E} \right|_{J, Q}, \quad \beta\Omega = \left. \frac{\partial S}{\partial J} \right|_{E, Q}, \quad \beta\Phi = \left. \frac{\partial S}{\partial Q} \right|_{E, J}. \quad (\text{A.2})$$

A.2 Canonical ensemble

The *canonical ensemble* is defined as the ensemble of possible states of a system in thermal equilibrium with a heat bath at some temperature T , and for given values of J, Q . Since

the system can exchange energy with the bath, the different states i will generically have different energy E_i . Each state is assigned a probability $P_i = \frac{1}{Z}e^{-\beta E_i}$, where $\beta = T^{-1}$ and $Z(\beta, J, Q)$ is the canonical partition function. The latter is defined as

$$Z(\beta, J, Q) = \sum_i e^{-\beta E_i}, \quad (\text{A.3})$$

where the sum is over all states i with assigned J, Q .¹⁸ It is also convenient to introduce the Helmholtz free energy F , given by

$$F(\beta, J, Q) = -\frac{1}{\beta} \log Z(\beta, J, Q). \quad (\text{A.4})$$

The average energy of the system is given by

$$E = \sum_i P_i E_i = -\frac{\partial \log Z}{\partial \beta} = \frac{\partial(\beta F)}{\partial \beta}. \quad (\text{A.5})$$

The entropy is given by

$$S = -\sum_i P_i \log P_i = \log Z + \beta E = -\frac{\partial F}{\partial T}. \quad (\text{A.6})$$

It follows that

$$F(T, J, Q) = E - TS, \quad \text{with} \quad S = -\left. \frac{\partial F}{\partial T} \right|_{J, Q} \quad (\text{A.7})$$

that is $F(T, J, Q)$ is the Legendre transform of the energy $E = E(S, J, Q)$, in which T replaces S as the independent variable. When β is used instead of the temperature, this relation can also be written as

$$\beta F = \beta E - S, \quad \text{with} \quad \beta = \left. \frac{\partial S}{\partial E} \right|_{J, Q} \quad (\text{A.8})$$

meaning that $(\beta F) = -\log Z$ is the Legendre transform of the entropy $S = \log Z_{\text{micro}}$, in which β replaces E as the independent variable.

A.3 Grand-canonical ensemble

In classical thermodynamics the *grand-canonical ensemble* is defined as the ensemble where the temperature and the chemical potential for the number of particles are specified (hence,

¹⁸Quantum mechanically, this reads $Z(\beta, J, Q) = \text{Tr} e^{-\beta H}$, where H is the Hamiltonian, and the sum is over its eigenstates with quantum number J, Q (the corresponding operators must commute with the Hamiltonian).

the states do not have fixed energy or number of particles). For us, the grand-canonical ensemble is the ensemble where none of the charges of the system is specified, while the temperature and all potentials are fixed.

In addition to the inverse temperature β , we introduce the angular velocity Ω and the electrostatic potential Φ . Each state is assigned a probability $P_i = \frac{1}{Z(\beta, \Omega, \Phi)} e^{-\beta(E_i - \Omega J_i - \Phi Q_i)}$, where

$$Z(\beta, \Omega, \Phi) = \sum_i e^{-\beta(E_i - \Omega J_i - \Phi Q_i)} \quad (\text{A.9})$$

is the grand partition function, and the sum is over all states at fixed β, Ω, Φ . It is useful to introduce the Gibbs free energy (or grand-potential) G , defined as

$$G(\beta, \Omega, \Phi) = -\frac{1}{\beta} \log Z(\beta, \Omega, \Phi). \quad (\text{A.10})$$

Then the average energy, average angular momentum, and average charge are given by

$$\begin{aligned} J &= \sum_i P_i J_i = \left. \frac{1}{\beta} \frac{\partial \log Z}{\partial \Omega} \right|_{\beta, \Phi} = - \left. \frac{\partial G}{\partial \Omega} \right|_{\beta, \Phi}, \\ Q &= \sum_i P_i Q_i = \left. \frac{1}{\beta} \frac{\partial \log Z}{\partial \Phi} \right|_{\beta, \Omega} = - \left. \frac{\partial G}{\partial \Phi} \right|_{\beta, \Omega}, \\ E &= \sum_i P_i E_i = - \left. \frac{\partial \log Z}{\partial \beta} \right|_{\Omega, \Phi} + \Omega J + \Phi Q = \left. \frac{\partial(\beta G)}{\partial \beta} \right|_{\Omega, \Phi} + \Omega J + \Phi Q, \end{aligned} \quad (\text{A.11})$$

and are of course functions of the temperature and chemical potentials. We thus see that (E, Q, J) are conjugate to (β, Φ, Ω) , respectively. The entropy is

$$S = - \sum_i P_i \log P_i = \log Z + \beta(E - \Omega J - \Phi Q) = - \left. \frac{\partial G}{\partial T} \right|_{\Omega, \Phi}. \quad (\text{A.12})$$

It follows that

$$G(T, \Omega, \Phi) = E - TS - \Omega J - \Phi Q, \quad (\text{A.13})$$

with

$$S = - \left. \frac{\partial G}{\partial T} \right|_{\Omega, \Phi}, \quad J = - \left. \frac{\partial G}{\partial \Omega} \right|_{T, \Phi}, \quad Q = - \left. \frac{\partial G}{\partial \Phi} \right|_{T, \Omega}, \quad (\text{A.14})$$

that is $G(T, \Omega, \Phi)$ is the Legendre transform of the energy $E(S, J, Q)$, in which T, Ω, Φ replace S, J, Q as independent variables. Using β instead of the temperature, we can also write

$$\beta G(\beta, \Omega, \Phi) = \beta E - (\beta \Omega) J - (\beta \Phi) Q - S(E, J, Q), \quad (\text{A.15})$$

with

$$\beta = \left. \frac{\partial S}{\partial E} \right|_{J, Q}, \quad \beta \Omega = \left. \frac{\partial S}{\partial J} \right|_{E, Q}, \quad \beta \Phi = \left. \frac{\partial S}{\partial Q} \right|_{E, J}, \quad (\text{A.16})$$

meaning that $(\beta G) = -\log Z$ is the Legendre transform of the entropy $S = \log Z_{\text{micro}}$, in which β, Ω, Φ replace E, J, Q as independent variables.

References

- [1] J. M. Bardeen, B. Carter and S. W. Hawking, *The Four laws of black hole mechanics*, *Commun. Math. Phys.* **31** (1973) 161–170.
- [2] D. Christodoulou, *Reversible and irreversible transformations in black hole physics*, *Phys. Rev. Lett.* **25** (1970) 1596–1597.
- [3] J. D. Bekenstein, *Black holes and the second law*, *Lett. Nuovo Cim.* **4** (1972) 737–740.
- [4] J. D. Bekenstein, *Black holes and entropy*, *Phys. Rev.* **D7** (1973) 2333–2346.
- [5] S. W. Hawking, *Gravitational radiation from colliding black holes*, *Phys. Rev. Lett.* **26** (1971) 1344–1346.
- [6] S. W. Hawking, *Particle Creation by Black Holes*, *Commun. Math. Phys.* **43** (1975) 199–220.
- [7] J. Polchinski, *The Black Hole Information Problem*, in *Proceedings, Theoretical Advanced Study Institute in Elementary Particle Physics: New Frontiers in Fields and Strings (TASI 2015): Boulder, CO, USA, June 1-26, 2015*, pp. 353–397, 2017. [1609.04036](https://arxiv.org/abs/1609.04036). DOI.
- [8] S. B. Giddings and W. M. Nelson, *Quantum emission from two-dimensional black holes*, *Phys. Rev.* **D46** (1992) 2486–2496, [[hep-th/9204072](https://arxiv.org/abs/hep-th/9204072)].
- [9] S. D. Mathur, *The Information paradox: A Pedagogical introduction*, *Class. Quant. Grav.* **26** (2009) 224001, [[0909.1038](https://arxiv.org/abs/0909.1038)].
- [10] G. W. Gibbons and S. W. Hawking, *Action Integrals and Partition Functions in Quantum Gravity*, *Phys. Rev.* **D15** (1977) 2752–2756.
- [11] S. W. Hawking, *The Path Integral Approach to Quantum Gravity*, in *General Relativity: An Einstein Centenary Survey*, pp. 746–789. 1980.
- [12] S. W. Hawking and D. N. Page, *Thermodynamics of Black Holes in anti-De Sitter Space*, *Commun. Math. Phys.* **87** (1983) 577.

- [13] E. Witten, *Anti-de Sitter space, thermal phase transition, and confinement in gauge theories*, *Adv. Theor. Math. Phys.* **2** (1998) 505–532, [[hep-th/9803131](#)].
- [14] R. M. Wald, *Black hole entropy is the Noether charge*, *Phys. Rev.* **D48** (1993) R3427–R3431, [[gr-qc/9307038](#)].
- [15] V. Iyer and R. M. Wald, *Some properties of Noether charge and a proposal for dynamical black hole entropy*, *Phys. Rev.* **D50** (1994) 846–864, [[gr-qc/9403028](#)].
- [16] T. Jacobson, G. Kang and R. C. Myers, *On black hole entropy*, *Phys. Rev.* **D49** (1994) 6587–6598, [[gr-qc/9312023](#)].
- [17] A. Sen, *Black hole entropy function and the attractor mechanism in higher derivative gravity*, *JHEP* **09** (2005) 038, [[hep-th/0506177](#)].
- [18] A. Sen, *Entropy Function and AdS(2) / CFT(1) Correspondence*, *JHEP* **11** (2008) 075, [[0805.0095](#)].
- [19] A. Sen, *Black Hole Entropy Function, Attractors and Precision Counting of Microstates*, *Gen. Rel. Grav.* **40** (2008) 2249–2431, [[0708.1270](#)].
- [20] A. Sen, *Quantum Entropy Function from AdS(2)/CFT(1) Correspondence*, *Int. J. Mod. Phys.* **A24** (2009) 4225–4244, [[0809.3304](#)].
- [21] A. Dabholkar and S. Nampuri, *Quantum black holes*, *Lect. Notes Phys.* **851** (2012) 165–232, [[1208.4814](#)].
- [22] H. K. Kunduri, J. Lucietti and H. S. Reall, *Near-horizon symmetries of extremal black holes*, *Class. Quant. Grav.* **24** (2007) 4169–4190, [[0705.4214](#)].
- [23] S. Ferrara, R. Kallosh and A. Strominger, *N=2 extremal black holes*, *Phys. Rev.* **D52** (1995) R5412–R5416, [[hep-th/9508072](#)].
- [24] A. Strominger, *Macroscopic entropy of N=2 extremal black holes*, *Phys. Lett.* **B383** (1996) 39–43, [[hep-th/9602111](#)].
- [25] S. Ferrara and R. Kallosh, *Supersymmetry and attractors*, *Phys. Rev.* **D54** (1996) 1514–1524, [[hep-th/9602136](#)].
- [26] N. Banerjee, S. Banerjee, R. K. Gupta, I. Mandal and A. Sen, *Supersymmetry, Localization and Quantum Entropy Function*, *JHEP* **02** (2010) 091, [[0905.2686](#)].

- [27] A. Dabholkar, J. Gomes and S. Murthy, *Quantum black holes, localization and the topological string*, *JHEP* **06** (2011) 019, [[1012.0265](#)].
- [28] A. Dabholkar, J. Gomes and S. Murthy, *Localization & Exact Holography*, *JHEP* **04** (2013) 062, [[1111.1161](#)].
- [29] A. Zaffaroni, *AdS black holes, holography and localization*, *Living Rev. Rel.* **23** (2020) 2, [[1902.07176](#)].