Chapter 2
Black Holes and Thermodynamics: The First Half Century

Daniel Grumiller, Robert McNees and Jakob Salzer

Abstract  Black hole thermodynamics emerged from the classical general relativistic laws of black hole mechanics, summarized by Bardeen–Carter–Hawking, together with the physical insights by Bekenstein about black hole entropy and the semi-classical derivation by Hawking of black hole evaporation. The black hole entropy law inspired the formulation of the holographic principle by ’t Hooft and Susskind, which is famously realized in the gauge/gravity correspondence by Maldacena, Gubser–Klebanov–Polaykov and Witten within string theory. Moreover, the microscopic derivation of black hole entropy, pioneered by Strominger–Vafa within string theory, often serves as a consistency check for putative theories of quantum gravity. In this book chapter we review these developments over five decades, starting in the 1960s.

Keywords  Black hole thermodynamics · History of black holes · Hawking radiation · Information loss · Holographic principle · Quantum gravity

2.1 Introduction and Prehistory

Introductory remarks. The history of black hole thermodynamics is intertwined with the history of quantum gravity. In the absence of experimental data capable of probing Planck scale physics the best we can do is to subject putative theories of quantum gravity to stringent consistency checks. Black hole thermodynamics provides a number of highly non-trivial consistency checks. Perhaps most famously, any theory of quantum gravity that fails to reproduce the Bekenstein–Hawking relation
between the black hole entropy \( S_{\text{BH}} \), the area of the event horizon \( A_h \), and Newton’s constant \( G \) would be regarded with a great amount of skepticism (see e.g [1]).

In addition to providing a template for the falsification of speculative models of quantum gravity, black hole thermodynamics has also sparked essential developments in the field of quantum gravity and remains a vital source of insight and new ideas. Discussions about information loss, the holographic principle, the microscopic origin of black hole entropy, gravity as an emergent phenomenon, and the more recent firewall paradox all have roots in black hole thermodynamics. Furthermore, it is an interesting subject in its own right, with unusual behavior of specific heat, a rich phenomenology, and remarkable phase transitions between different spacetimes.

In this review we summarize the development of black hole thermodynamics chronologically, except when the narrative demands deviations from a strictly historical account. While we have tried to be comprehensive, our coverage is limited by a number of factors, not the least of which is our own knowledge of the literature on the subject. Each of the following five sections describes a decade, beginning with the discovery of the Kerr solution in 1963 [2]. In our concluding section we look forward to future developments. But before starting we comment on some early insights that had the potential to impact the way we view the result (2.1).

**Prehistory.** If the history of black hole thermodynamics begins with the papers of Bekenstein [3] and Bardeen et al. [4], then the prehistory of the subject stretches back nearly forty additional years to the work of Tolman, Oppenheimer, and Volkoff in the 1930s [5–7]. These authors considered the conditions for a ‘star’—a spherically symmetric, self-gravitating object composed of a perfect fluid with a linear equation of state—to be in hydrostatic equilibrium. Later, in the 1960s, Zel’dovich showed that linear equations of state besides the familiar \( p = 0 \) (dust) and \( p = \rho/3 \) (radiation) are consistent with relativity [8]. He established the bound \( p \leq \rho \), with \( p = \rho \) representing a causal limit where the fluid’s speed of sound is equal to the speed of light. A few years after that, Bondi considered massive spheres composed of such fluids and included the case \( p = \rho \) in his analysis [9].

The self-gravitating, spherically symmetric perfect fluids considered by these and other authors possess interesting thermodynamic properties. In particular, the entropy of such objects (which are always outside their Schwarzschild radius) is not extensive in the usual sense. For example, a configuration composed of radiation has an entropy that scales with the size of the system as \( S(R) \sim R^{3/2} \), and a configuration with the ultra-relativistic equation of state \( p = \rho \) has an entropy \( S(R) \sim R^2 \) that scales like the area. But these results do not appear in the early literature (at least, not prominently) because there was no compelling reason to scrutinize the relationship between the entropy and size of a gravitating system before the 1970s. It was not until the 1980s, well after the initial work of Bekenstein and Hawking, that Wald, Sorkin, and Zhang studied the entropy of self-gravitating perfect fluids with \( p = \rho/3 \) [10]. They showed that the conditions for hydrostatic equilibrium—the same conditions
set out by Tolman, Oppenheimer, and Volkoff—are at least local extrema of the entropy. With reasonable physical assumptions these objects quite easily satisfy the Bekenstein bound, \( S \leq 2\pi k_B E / (\hbar c) \), where \( R \) and \( E \) are the object’s size and energy, respectively.

The area law (2.1) is often presented as a surprising deviation from the volume scaling of the entropy in a non-gravitating system. But the early work described above suggests, without invoking anything as extreme as a black hole, that this is something we should expect from General Relativity. Even a somewhat mundane system like a sufficiently massive ball of radiation has an entropy that is not proportional to its volume. The surprising thing about the area law is not that the entropy of the system grows much more slowly than a volume. Rather, it is that the entropy of a black hole seems to saturate, at least parametrically, an upper bound on the growth of entropy with the size of a gravitating system. Such a bound, which follows from causality, could have been conjectured several years before the work of Bekenstein and Hawking.

### 2.2 1963–1973

**Black hole solutions and the uniqueness theorem.** After the first black hole solutions were found in immediate consequence to the publication of Einstein’s equations, it took almost 50 years for the next exact black hole solution to be discovered. The Kerr solution [2] describes a rotating black hole of mass \( M \) and angular momentum \( J = aM \)

\[
ds^2 = -\left(1 - \frac{2Mr}{\rho^2}\right) dt^2 - \frac{4Mr a \sin^2 \theta}{\rho^2} dt \, d\phi + \left(r^2 + a^2 + \frac{2Mra^2 \sin^2 \theta}{\rho^2}\right) \sin^2 \theta \, d\phi^2 + \frac{\rho^2}{r^2 - 2Mr + a^2} dr^2 + \rho^2 \, d\theta^2 \quad \text{with} \quad \rho^2 := r^2 + a^2 \cos^2 \theta.
\]

(2.2)

Only 2 years later this solution was extended to include charged rotating black holes [11]. These black hole solutions exhibit the remarkable property that they are parameterized in terms of only three quantities as measured from infinity: mass, angular momentum, charge. It was therefore natural to ask whether this was the case for all black hole solutions.

Building on earlier work concerning the persistence of the horizon under asymmetric perturbations [12, 13], Israel proved that—assuming some regularity conditions—the Schwarzschild solution is the only static, asymptotically flat vacuum spacetime that exhibits a regular horizon [14]. Later, this proof was generalized to static asymptotically flat electrovac spacetimes, now with the Reissner–Nordström black hole as the only admissible spacetime [15]. In the case of axisymmetric stationary black holes Carter was later able to show that these spacetimes fall into discrete sets of continuous families, each of them depending on one or two independent parameters, with the Kerr solutions as the unique family to allow vanishing angular momentum [16]. The key point of Carter’s proof is the observation that Einstein’s equations for
an axisymmetric spacetime can be reduced to a two-dimensional boundary value problem. Building on this, Robinson showed that in fact only the Kerr family exists, thus establishing the uniqueness of the Kerr black hole [17]. Similar results concerning the classification and uniqueness of charged axisymmetric stationary black holes were worked out independently by Mazur [18], Bunting [19] and more recently by Chruściel and Costa [20]. However, due to different initial hypotheses in the statement of the theorem and some technical gaps, the uniqueness theorem is still extensively studied (cf. [21] for a review).

Referring to these results, John Wheeler coined the expression “black holes have no hair” [22], i.e. black holes can be described entirely by a small amount of quantities measured from infinity. The no-hair conjecture thus suggests a resemblance of black holes to systems in thermodynamic equilibrium, whose macroscopic state is parameterized by a small number of macroscopic variables.

**Penrose process and superradiant scattering.** Another similarity between black holes and thermodynamical systems was revealed with Penrose’s suggestion that energy can be extracted from a rotating black hole [23]. The Penrose process relies on the presence of an ergosphere in Kerr spacetime. In this region the Killing field $\xi^a$ that asymptotically corresponds to time translation is spacelike. Consequently, the energy $E = -\rho_a \xi^a$ of a particle of 4-momentum $\rho^a$ need not be positive. In the Penrose process a particle with positive energy $E_0$ is released from infinity. In the ergosphere the particle breaks up in such a way that one fragment has negative energy $E_1$ whereas the other has positive energy $E_2 = E_0 - E_1 > E_0$. If the latter returns back to infinity on a geodesic one has effectively gained the energy $|E_1|$. The negative energy particle falls into the black hole and therefore reduces its mass. Thus, energy is indeed extracted from the black hole. Angular momentum $j_2$ and energy of the particle falling into the black hole have to obey the inequality $j_2 \leq \frac{E_2}{\Omega_H}$, where $\Omega_H$ is the angular velocity of the black hole. Therefore, the change in the black hole’s mass and angular momentum $\delta M$ and $\delta J$, respectively, are related by $\delta M \geq \Omega_H \delta J$. This equation can be rewritten in a form that bears a clear resemblance to the second law of thermodynamics [24]

$$\delta M_{irr} \geq 0,$$

(2.3)

where $M_{irr}^2 = \frac{1}{2} \left(M^2 + \sqrt{M^4 - J^2}\right)$ is the irreducible mass. Expressed in terms of irreducible mass and angular momentum, the mass of the black hole reads

$$M^2 = M_{irr}^2 + \frac{J^2}{4M_{irr}^2} \geq M_{irr}^2.$$

(2.4)

The maximum amount of energy that can be extracted from a black hole with initial mass $M_0$ and angular momentum $J_0$ is therefore $\Delta M = M_0 - M_{irr}(M_0, J_0)$, which is maximized for an extremal black hole, i.e. $J_0 = M_0^2$, with an efficiency of 0.29.

A generalization to charged rotating black holes yields the Christodoulou–Ruffini mass formula
\[ M^2 = \left( M_{irr} + \frac{Q^2}{4M_{irr}} \right)^2 + \frac{J^2}{4M_{irr}^2}, \quad (2.5) \]

which pushes the efficiency of the Penrose process up to 0.5 [25].

The fact that a Penrose process cannot reduce the irreducible mass of a black hole is a particular consequence of Hawking’s area theorem, discussed below.

The Penrose process has a corresponding phenomenon in wave scattering on a stationary axisymmetric black hole background known as superradiant scattering [26–28]. Similar effects were already studied in [29, 30] where scalar waves incident on a rotating cylinder were examined. For a qualitative understanding of superradiant scattering consider the scalar wave equation \( \nabla^a \nabla_a \Phi = 0 \) on a Kerr background. It was shown in [31] by studying the Hamilton–Jacobi equation for a test particle that this equation is separable, therefore \( \Phi \) can be written as:

\[ \Phi = e^{i(m\phi-\omega t)} R(r) P(\theta) \]

where \( P(\theta) \) is a spheroidal harmonic. The solutions for \( R(r) \) were studied in detail in [32]. Suitable boundary conditions for \( \Phi \) read

\[ \Phi(r) = \begin{cases} e^{-i(\omega-m\Omega)r_+} & r \rightarrow r_+ \\ A_{out}(\omega)e^{ior_+} + A_{in}(\omega)e^{-ior_+} & r \rightarrow \infty \end{cases} \quad (2.6) \]

where \( r_+ \) denotes the tortoise coordinate for the Kerr spacetime. The choice of boundary condition at the horizon \( r \rightarrow r_+ \) is motivated by the requirement that physical observers at the horizon should see exclusively ingoing waves. The Wronskian determinant for this solution and its complex conjugate evaluated in both limits leads to

\[ |R|^2 = 1 - \left( 1 - \frac{m\Omega_H}{\omega} \right) |T|^2. \quad (2.7) \]

Therefore, superradiance is observed for \( \omega < m\Omega_H \). Interestingly, the amplification of the incoming amplitude depends on the spin of the incident wave [33, 34]: 0.003 for a scalar wave, 0.044 for an electromagnetic field and 1.38 for gravitational waves. Half-integer fields do not appear, as fermions show no superradiant scattering behavior. This can be understood from the exclusion principle which allows only one particle in each outgoing mode and thus prevents an enhancement of the scattered wave [35, 36].

The occurrence of superradiant scattering in quantum mechanics is well known from the Klein paradox [37–39]. The Klein paradox describes the quantum effect that a wave incident on a step potential is reflected with a coefficient \( |R| > 1 \) for a particular relation between potential height and energy of the incident wave. This effect is attributed to pair creation in the strong electric field near the potential step. Therefore, the presence of superradiant scattering in a black hole background suggests the occurrence of particle creation as was already noted in [28–30, 33] and later famously shown by Hawking [40] (cf. next section).

**The area theorem.** The above mechanisms of energy extraction are closely related to the important area theorem. The area theorem and the four laws of black hole
mechanics rely on a couple of earlier theorems, which are described in the following with no intention of mathematical rigor (cf. the standard reference \[41\] for details).

The rigidity theorem shows under suitable assumptions that the event horizon of a stationary black hole is a Killing horizon. This result can be proven in two independent ways. Carter showed that the horizon of a static black hole is normal to the static Killing vector $\xi^a$, and the horizon of a stationary black hole is normal to the linear combination $\chi^a = \xi^a + \Omega_H \phi^a$ under the assumption of $t - \phi$ orthogonality \[42\]. Here $\Omega_H$ denotes the angular velocity of the horizon and $\phi^a$ is the Killing vector generating the axial symmetry. In the second proof, Einstein’s equations are assumed in order to show that the event horizon of every stationary black hole in vacuum or electrovacuum is a Killing horizon \[41\].

The Penrose theorem proven in \[43\] states that the null geodesics generating the horizon may have past end points but no end points in the future. In particular, no caustics of the generators can occur when extended into the future. A consequence of this theorem is that black holes cannot bifurcate or vanish \[41\].

The focusing theorem follows from the Raychaudhuri equation for lightlike congruences. It states that, given a positive convergence at any point of the congruence, the cross-section of the beam vanishes in a finite distance provided the weak energy condition and Einstein’s equations hold.

The area theorem follows from the two latter statements: If the lightlike generators of the horizon had a positive convergence at any point, a caustic would occur in finite distance which is forbidden by Penrose’s theorem. Therefore, the area of the horizon cannot decrease \[44, 45\]. The only possibility to evade this conclusion is the presence of a naked singularity, i.e. a singularity not shielded by a horizon. Thus, the presence of such singularities must be excluded by adopting the cosmic censorship conjecture \[23\]. In summary, the prerequisites of the area theorem put strong restrictions on both causal structure (cosmic censorship conjecture) and matter (weak energy condition) in spacetime. In particular, the latter is in general not met when quantum theory is taken into account \[46\].

The area theorem provides an explanation for the bound on the Penrose process \[2.3\], since $M_{rr}$ is proportional to the area of the Kerr black hole \[44, 45\]. Similarly, the need for superradiant scattering of waves as described above and the existence of a spin-spin interaction between a Kerr black hole and a spinning particle can both be seen just from the area law \[46, 47\]. The argument for superradiant scattering from the area theorem breaks down for fermions since the respective energy-momentum tensor does not obey the weak energy condition \[35, 36\].

**The four laws of black hole mechanics.** The work of this decade culminated in the famous four laws of black hole mechanics by Bardeen et al. \[4\]. These laws, which show a remarkable similarity to the laws of thermodynamics, are stated in the following.

- The zeroth law of black hole mechanics states that the surface gravity $\kappa$ is constant on the horizon of a black hole. The proof of this result as given in \[4\] requires the dominant energy condition to hold and the use of Einstein’s equations. Similar results were already obtained in \[16, 45\]. Another proof was given in \[42\] by using
the assumption of $t - \phi$ orthogonality [42]. The zeroth law suggests a similarity between $\kappa$ and the temperature of a body in thermal equilibrium.

- The first law establishes a relation between changes in the mass $M$, horizon area $A$, angular momentum $J_H$, and charge $Q_H$ if the black hole is perturbed.

$$\delta M = \frac{\kappa}{8\pi} \delta A + \Omega_H \delta J_H + \Phi_H \delta Q$$  \hspace{1cm} (2.8)

In [4] the first law was derived from a generalized version of the Smarr mass formula [48]. The first law bears a clear resemblance to the first law of thermodynamics with $\kappa$ as temperature, the horizon area $A$ taking the place of entropy, and the mass $M$ taking the role of energy. As pointed out in [49], in fact two different versions of the first law exist: an equilibrium version, wherein one compares the parameters of neighboring equilibrium solutions, and a physical process version in which the parameters of the black hole are changed, e.g. by dropping in matter, and analyzing the change in the parameters after the black hole has settled down. These two independent versions yield the same result.

- The second law of black hole mechanics is Hawking’s area theorem:

$$\delta A \geq 0.$$  \hspace{1cm} (2.9)

Here, the analogy between horizon area $A$ and entropy becomes evident.

- The third law states that the surface gravity of a black hole cannot be reduced to zero in a finite number of processes. This formulation is an analogue of the Nernst unattainability principle [50]. The Planck formulation of the third law of thermodynamics does not hold in black hole mechanics, as the horizon area of an extremal black hole is finite despite vanishing surface gravity. It follows from the third law that non-extremal black holes cannot be made extremal in a finite number of steps. A proof for the third law was presented later in [51].

The close mathematical and physical analogy between the four laws of black hole mechanics and the laws of thermodynamics is remarkable. Nonetheless, it appears to be a mere analogy in classical general relativity. Classical black holes do not have temperature since they cannot radiate, and entropy is a dimensionless quantity in contrast to the horizon area that has a dimension of length squared. It is only when quantum theory is taken into account that the analogy becomes an identity.

2.3 1973–1983

**Bekenstein–Hawking entropy.** In 1971, when Wheeler proposed the now famous gedankenexperiment of pouring a hot cup of tea into a black hole, he was questioning whether black holes violate the second law of thermodynamics. Another possible violation of the second law of thermodynamics by classical black holes was put forward by Geroch: A box of matter with mass $m$ is lowered close to the horizon
of a black hole from infinity where its energy (as measured from infinity) is nearly zero, thus providing an amount $m$ of work. At the horizon the box radiates away an amount $\delta m$ of its mass and is hauled back to infinity, requiring an amount $m - \delta m$ of work. In this process an amount $\delta m$ of heat is transformed entirely into work thereby violating the second law of thermodynamics.

These evident violations of the second law of thermodynamics and Hawking’s area theorem led Bekenstein to propose an entropy for black holes that is proportional to horizon area measured in units of Planck area with a coefficient $\eta$ of order one [3, 52]:

$$S_{BH} = \eta \frac{k_B c^3 A}{\hbar G}.$$  \hspace{1cm} (2.10)

The second law of thermodynamics is then replaced by a generalized second law of thermodynamics which states that the change in the sum of matter entropy and black hole entropy is strictly nonnegative. Bekenstein showed that this generalized second law resolves the problems associated with the Geroch process [52]. Furthermore, he tested the law for the cases of a harmonic oscillator enclosed in a spherical box and infalling radiation.

**Hawking radiation.** The relation between black hole entropy and horizon area together with the first law of black hole mechanics indicates that black holes do have a temperature that should be proportional to $\kappa$. If the black hole is immersed in black body radiation of lower temperature then the generalized second law is violated, unless the black hole also emits radiation. Therefore, spontaneous particle creation is needed to prevent a violation of the generalized second law. Eventually, Hawking showed that black holes spontaneously emit radiation characteristic of a black body at temperature

$$T_H = \frac{\hbar \kappa}{2\pi c k_B},$$  \hspace{1cm} (2.11)

thus establishing also further evidence for the validity of the generalized second law [36, 40]. The coefficient $\eta$ in (2.10) is fixed to $\eta = \frac{1}{4}$ as can be seen from the first law of black hole mechanics (2.8). Consequently, black holes can be treated as thermodynamic systems, and the four laws of black hole mechanics cease to be mere analogies, but describe black holes as thermodynamic systems.

In the original derivation of Hawking radiation a massless scalar field is studied in the background metric of gravitational collapse. The scalar field yields a decomposition in terms of a complete set of solutions both on lightlike past infinite, $I^-$, and on the union of the horizon and lightlike future infinity, $I^+$. Both sets of solutions contain positive frequency modes with respect to the appropriate affine parameters on $I^+$ and $I^-$. The different decompositions of the field induce a Bogoliubov transformation on the two sets of creation and annihilation operators on $I^+$ and $I^-$. Therefore, the vacuum state with respect to the operators for the ingoing particles
yields a nonzero expectation value for the number operator for an observer in $\mathcal{I}^+$.\footnote{In quantum information language the vacuum quantum state in a black hole space-time for each mode is a two-mode squeezed vacuum, similar to what happens for primordial density fluctuations in cosmology [53].} This is the Hawking radiation. The particle number measured at $\mathcal{I}^+$ of a particular mode is given by [36]  
\[ \langle n_J \rangle = \frac{\Gamma_J}{\exp\left[(2\pi \bar{\omega}_J)/\kappa\right]} \pm 1, \]  
(2.12)

where the index $J$ denotes collectively frequency $\omega$, angular momentum $l$, azimuthal quantum number $m$, sign of the charge, and spin. The upper sign is for fermions and the lower sign for bosons. Here the quantity $\bar{\omega}$ is defined as $\bar{\omega}_J = \omega_J - m_j \Omega_H - q \Phi_H$, and $\Gamma_J$ denotes the fraction of the incident radiation that enters the collapsing body, i.e. $\Gamma_J = 1 - |R_J|^2$. Expression (2.12) is precisely the result expected for a black body with temperature $\kappa/(2\pi)$ and greybody factor $\Gamma_J$. In fact, Hawking radiation is completely identical to black body radiation, since the density matrices for Hawking radiation and black body radiation coincide [54–56]. Furthermore, it was shown that black holes behave like black bodies even in the presence of incoming radiation. Expressions for the probability of emission of $k$ particles when $m$ particles have arrived, $P(k|m)$, and the Einstein coefficients for induced emission, spontaneous emission and absorptions were obtained in [57, 58]. The derivation of Hawking was repeated subsequently in various approaches and generalizations (cf. [59–63]).

The Hawking effect is often described heuristically as Schwinger pair creation in the gravitational field of a black hole, where the negative energy particle drops into the black hole and the other particle escapes to infinity [36]. A derivation of the Hawking effect that closely resembles this picture of a tunneling process was presented in [64].

**Hawking radiation from anomalies.** Particularly striking is the connection between Hawking radiation and anomalies of the stress-energy tensor. If restricted to the s-wave sector, Hawking radiation can be studied in an effectively two dimensional spacetime. In this geometry, Hawking radiation can be shown to arise from the trace anomaly of the energy-momentum tensor for a massless field, by requiring finiteness of $T_{\mu\nu}$ at the horizon for a geodesic observer [65] (cf. e.g. [66] for a general discussion, and [67]).

More recently, it was shown that Hawking radiation is necessary for the cancellation of gravitational anomalies [68]—i.e. non-conservation of $T_{\mu\nu}$—in Schwarzschild spacetimes of any dimension. A gravitational anomaly occurs if one assumes that modes propagating along the horizon can be integrated out, so that $T_{\mu\nu}$ is regular on the horizon. Thus, the resulting theory is effectively chiral near the horizon and acquires a gravitational anomaly, which is removed by Hawking radiation. This method can be generalized to charged and rotating black holes [69, 70].

**Euclidean path integral.** Due to its intimate connection with the partition function, the Euclidean path integral formalism of quantum gravity is widely used when studying black hole thermodynamics.
The singularities encountered in black hole spacetimes can be avoided by a Wick rotation into the Euclidean sector. This requires a periodic Euclidean time with periodicity of inverse temperature. In general, the Euclidean path integral does not converge due to the presence of the conformal mode [71]. However, in the semi-classical approximation to the partition function, i.e. expanding the path integral around solutions of the classical equations of motion, the above results for entropy and temperature of black holes are recovered [72, 73].

In the case of flat spacetime at non-zero temperature studied in [74], a sum over the Schwarzschild instanton in the path integral leads to a non-zero probability for the decay of flat space into a black hole. Gravitational instantons and their thermodynamic properties were studied and classified in [75–77].

**Modes of black hole decay.** It is seen from (2.12) that the emission of particles with charge of the same sign as the charge of the black hole is enhanced. Thus, the charge of the black hole is radiated away [78–82]. The resulting current is proportional to the particle number (2.12) times the charge of the emitted particle. An estimation of the discharge rate yields that the timescale over which the discharge occurs is in general much shorter than the relevant timescale for formation of the black hole [79], provided that \( Q/M \geq Mm_e^2/e \). Thus, only very large black holes show a significant charge \( Q \). All other black holes show only random charge fluctuations of order \((\hbar c)^{1/2}\) after sufficiently long time [81, 82].

For fixed angular momentum \( l \) the emission of particles with positive azimuthal quantum number \( m \) is enhanced, and the black hole loses angular momentum. When radiation of massless particles only is considered, the black hole loses angular momentum considerably faster than mass [83]. Curiously, the emission of neutrinos shows parity violation: antineutrinos are emitted preferentially parallel to angular momentum whereas more neutrinos are emitted in the opposite direction [84–86].

Radiation of the black hole mass occurs over a timescale \( \tau \propto G^2 M^3_0/(\hbar c^4) \), which exceeds the age of the present universe unless the black hole is sufficiently light, \( M_0 \leq 5 \times 10^{11} \) kg. The species of the emitted particles changes with the mass of the black hole: black holes emit massless particles only as long as \( M \geq 10^{14} \) kg at which point electron-positron emission starts; the onset for emission of heavier particles lies at \( M \approx 10^{11} \) kg [81–83]. Consequences of black hole evaporation for unitarity are discussed below.

**Unruh effect.** The Unruh effect describes the detection of vacuum fluctuations of the Minkowski vacuum as thermal radiation by a constantly accelerated observer, i.e. an observer in Rindler spacetime [61, 87–89]. The Minkowski vacuum—the vacuum for an observer measuring time along the Killing vector \( \partial_t \)—can be represented as the sum

\[
|0\rangle = \sum_n \exp(-2\pi a^{-1} \omega_n)|n\rangle_L \times |n\rangle_R. \tag{2.13}
\]

where \( |n\rangle_L(|n\rangle_R) \) are states with energy \( \omega_n \) measured by an observer moving with acceleration \( a \) along the respective Killing vectors in the left (right) wedge of Rindler
The Minkowski vacuum thus contains correlations between states in different wedges of Rindler spacetime and is regarded as a thermal bath of temperature

$$T_U = \frac{\hbar a}{2\pi c k_B}$$

(2.14)

by the accelerating observer. Although the original derivation was given for free fields, the validity of the Unruh effect for interacting fields is a consequence of general results obtained in axiomatic quantum field theory [91, 92], as first recognized in [93]. The Unruh effect indicates that already in flat spacetime the notion of particles is observer dependent. Other seemingly paradoxical aspects of the Unruh effect are covered in [94]. Recent developments and issues regarding experimental detection are reviewed in [95].

The Unruh effect is also invoked to prevent a violation of the generalized second law in the following form: A box with given energy and entropy is released from infinity and its content dropped into the black hole. The energy gain of the black hole can be made arbitrarily small by dropping the box from a point close to the horizon. The horizon area might not increase enough to compensate the loss of entropy, thus violating the generalized second law. In [96], a universal upper bound on the ratio entropy to energy was proposed, which would prevent such violations of the generalized second law. On the other hand, it was argued that the box would feel an effective buoyancy force near the black hole originating from the acceleration radiation. This buoyancy force guarantees a lower bound on the energy gain of the black hole, thus saving the generalized second law without the need for an entropy bound [97, 98].

The similarities between Hawking and Unruh effect are due to the similar horizon structure: any non-extremal Killing horizon looks like a Rindler horizon in the near-horizon approximation. Depending on the choice of boundary conditions different vacua exist, which are suitable for different physical applications. The Unruh vacuum fixes boundary conditions on the past horizon $H^-$ and $\mathcal{I}^-$ [88]. This state is analogous to the original treatment of black hole evaporation by Hawking. For the Hartle–Hawking vacuum one defines boundary conditions on both future and past horizon $H^+$ and $H^-$ [99, 100], which describes a black hole in equilibrium with incoming radiation, and is therefore the relevant state for the curved spacetime generalization of the Unruh effect. This state does not exist for Kerr black holes [101]. The Boulware vacuum sets boundary condition on $\mathcal{I}^+$ and $\mathcal{I}^-$ and describes a state with no radiation [59], but is singular on past and future horizon and therefore of little physical significance.

The transplanckian problem and black hole analogue systems. Since the direct experimental verification of black hole thermodynamics effects is (and most likely will remain) out of reach, analog systems have been proposed in which the Hawking effect could be studied. One of the first proposed systems concerns sound waves
in a convergent fluid flow [102]. The linearized equations of motion correspond to the equations for a massless scalar field in a background metric that can be brought in a Schwarzschild-like form, thus producing a sonic black hole with the speed of light replaced by the speed of sound. Quantization of the scalar field in this background leads to the emission of sound waves in a thermal spectrum at the sonic horizon, the temperature of which is given by a quantity analogous to the Hawking temperature. Albeit very small, this quantity should be measurable in principle. The field of analogue gravity has grown rapidly in the last decades; the interested reader is referred to [103, 104] and references therein.

Black hole analogue systems play an important role in the study of the transplanckian puzzle. A Hawking mode of frequency $\omega$ measured at infinity that was emitted a time $t$ after formation of the black hole stems from a fluctuation of frequency $\omega \exp(\kappa t)$. This means that modes emitted a sufficiently long time after the formation of the black hole originated from modes beyond the Planck energy, where the theory can no longer be trusted. This raises the question if the Hawking effect depends on the details of a transplanckian theory. Certainly, Lorentz invariance would guarantee the validity of the derivation, but it is a logical possibility (though one that is highly-constrained by observations from the Fermi Large Area Telescope [105]) that Lorentz invariance is broken at arbitrarily high energies, see [106] for a discussion. A viable option, at least for analogue systems, is the study of Hawking radiation with a modified dispersion relation at high frequencies. Since this situation is similar to the study of black hole analog systems in fluid mechanics, where the theory breaks down at wavelengths comparable to the atomic scale, these systems are used in the study of the transplanckian problem. A particular example was presented in [107], where it was shown that Hawking radiation occurs despite a change in the dispersion relation at high frequencies.

The transplanckian problem was far from being settled in that decade, but it seems that Hawking radiation is robust enough to persist, even if the theory is modified at ultrahigh energies like in analogue systems [108–113].

**Black hole evaporation and information loss.** As pointed out above, black holes evaporate due to Hawking radiation on a timescale $\tau \propto G^2 M_0^3/(\hbar c^4)$, which is of order $10^{70}$s for a solar mass black hole. Although this amount of time is enormous already for solar mass black holes, the very fact that black holes evaporate reveals the deep conceptual problem of information loss, first raised in [56]. At the classical level, the no-hair theorem implies that the large amount of data needed to describe the precollapse geometry is reduced to a small number of quantities that describe the black hole. The remaining information of the precollapse geometry is not accessible to the outside observer, but in principle can be thought of as residing in the black hole. The real paradox rears its head when Hawking radiation is taken into account. Consider an initial pure state that describes an object falling into the black hole. The Hawking radiation emitted by the black hole is in a mixed state due to correlations between states outside the horizon and states inside the black hole, but after some time the black hole has evaporated completely, and one is left only with the mixed state of Hawking radiation. The evolution from the initial pure state to perfectly thermal
Hawking radiation is therefore not unitary and information appears to be lost in the process. This is in contrast to ordinary physical systems like a star or a burning lump of coal, where the emissions contain correlations that would in principle allow one to reconstruct the initial state. It was not clear at the time if this might also be a viable explanation for an evaporating black hole, mostly due to the lack of a sufficiently detailed theory of quantum gravity.

2.4 1983–1993

The results of the previous decade revealed several problems—the information paradox, the universality of the area law, and the nature of the states underlying the Bekenstein-Hawking entropy—that became the focus of research during the period 1983–1993. Many researchers turned their attention towards lower-dimensional models, where theories are more tractable but still suffer from conceptual issues such as the information paradox. At the same time, investigations into a diverse array of gravitational theories revealed certain universal features of black hole thermodynamics and led to the first early successes in a state-counting approach to explaining black hole entropy. Before delving into lower-dimensional gravity we state some of the main conclusions that were reached from its study.

What to do with information loss? The information loss problem is of conceptual rather than technical nature. Like other conceptual issues in classical and quantum gravity, it arises independently from the spacetime dimension. Therefore, a useful strategy is to consider lower-dimensional models of gravity where the technical problems become more manageable, conceptual issues can be addressed and, ideally, resolved. See [114] for a textbook on lower-dimensional gravity from 1988. Particularly the CGHS model of string-inspired 2-dimensional dilaton gravity with matter [115] (see below) inspired numerous investigations of evaporating black holes in two dimensions, such as the one by Russo et al. [116]. Exact solubility (even in the presence of quantum effects) is a key feature of the RST model, which allows to address the endpoint of Hawking evaporation. Depending on the energy flux of the infalling matter either no horizon forms or an apparent horizon does form and eventually evaporates to a naked singularity, which requires the imposition of suitable boundary conditions, for which a natural choice exists in this model. Most importantly, the whole process is described in a unitary way, so that all information is recovered in this case.

Black hole complementarity. Based on studies of 2-dimensional dilaton gravity models, Susskind, Thorlacius and Uglum advocated the “black hole complementarity” principle [117] (which was formulated independently in [118]). The essence of this principle is captured by four postulates (three of which were spelled out explicitly in [117], which we quote verbatim): 1. The process of formation and evaporation of a black hole, as viewed by a distant observer, can be described entirely within the
context of standard quantum theory. In particular, there exists a unitary S-matrix which describes the evolution from infalling matter to outgoing Hawking-like radiation. 2. Outside the stretched horizon\(^3\) of a massive black hole, physics can be described to good approximation by a set of semi-classical field equations. 3. To a distant observer, a black hole appears to be a quantum system with discrete energy levels. The dimension of the subspace of states describing a black hole of mass \(M\) is the exponential of the Bekenstein entropy (2.1). 4. A freely falling observer experiences nothing extraordinary when entering the black hole.

The attribute “complementarity” refers to the fact that the outside observer detects a membrane-like structure near the black hole horizon where information is stored, while the infalling observer sees no membrane at the horizon. The reason why these mutually exclusive viewpoints do not necessarily generate a contradiction is because there should not exist any “super-observer” that simultaneously has access to both viewpoints.

**Lower-dimensional gravity.** The lowest spacetime dimension that makes sense to consider is \(1 + 1\), since this is the lowest dimension where the notions of black holes, causal structure and curvature exist. If additionally the existence of graviton excitations (at least off-shell) is required then the lowest dimension one can consider is \(2 + 1\), since this is the lowest dimension where linearized perturbations of the metric \(h_{\mu\nu}\) have a transverse-traceless part, \(h_{\mu\nu} = h^{TT}_{\mu\nu} + \nabla_{(\mu} \xi_{\nu)} + \frac{1}{3} h g_{\mu\nu}\). Moreover, \(2 + 1\) is the lowest dimension where the notion of the area of the event horizon is meaningful (in \(1 + 1\) dimensions this ‘area’ is just a point). For these reasons, the main focus in lower-dimensional gravity is on \(1 + 1\) and \(2 + 1\) dimensional models, depending on the scope of the model.

**Dilaton gravity in two dimensions.** In two dimensions there are various ways to motivate which kind of gravity model one should consider. The theory not to consider is Einstein gravity, since there are no meaningful Einstein equations in two dimensions (the 2-dimensional Einstein tensor vanishes trivially for any metric). Instead, there are (at least) five different ways to end up with the same class of models, namely 2-dimensional dilaton gravity. Its bulk action

\[
I = \frac{1}{16\pi G} \int d^2 x \sqrt{-g} \left( XR - U(X)(\partial X)^2 - 2V(X) \right) \tag{2.15}
\]

depends on two free functions, \(U(X)\) and \(V(X)\), of the dilaton field \(X\). We summarize briefly five different ways to end up with an action of type (2.15).

1. **Gravity as gauge theory.** Jackiw [120] and Teitelboim [121] considered a 2-dimensional gravity model with constant curvature, which can be formulated as a non-abelian BF-theory with gauge group \(SO(2, 1)\) [122, 123]. The generators \(P_a\) and \(J\) are interpreted as translation and boost generators, respectively. They obey the algebra \([P_a, P_b] = \Lambda \varepsilon_{ab} J\) and \([P_a, J] = \varepsilon_{ab} P_b\), where \(\Lambda\) is a

\(^3\) The stretched horizon (or the earlier “brick wall” [119]) is also discussed in these papers and captures the membrane description of a black hole suitable for a distant observer.
parameter that sets the scale of curvature (one could call it ‘cosmological constant’). The \( so(2, 1) \) connection \( A = e^a P_a + \omega J \) decomposes into zweibein \( e^a \) and (dualized) connection \( \omega = \frac{1}{2} \omega^{ab} \varepsilon_{ab} \). Its non-abelian field strength \( F \) is then coupled linearly to co-adjoint scalars in the BF-action, which reads explicitly

\[
I \sim \int \left( X_a \left( de^a + \varepsilon^{ab} \omega e^b \right) + X d\omega + \varepsilon_{ab} e^a \wedge e^b \Lambda X \right). \tag{2.16}
\]

Integrating out the auxiliary field \( X_a \) establishes the constraint of vanishing torsion, which allows to eliminate also the spin-connection \( \omega \) and to convert the first order action (2.16) into the second order action (2.15) with \( U(X) = 0 \) and \( V(X) = \Lambda X \). A similar BF-type of construction was provided by Cangemi and Jackiw [124] for a string inspired model discussed below. The gauge theoretic formulation for arbitrary dilaton gravity theories was provided by Ikeda and Izawa [125, 126] and by Schaller and Strobl [127], dubbed “Poisson-\( \sigma \) model”.

2. **Dimensional reduction.** Assuming spherical symmetry in \( D \) spacetime dimensions leads to a line-element in adapted coordinates that depends on a 2-dimensional metric and a scalar field, \( ds^2 = g_{\alpha\beta} dx^\alpha dx^\beta + X^{1/(D-2)} d\Omega_{SD-2}^2 \), where \( d\Omega_{SD-2}^2 \) denotes the line-element of the round \( (D-2) \)-sphere [128–132]. Inserting this ansatz into the \( D \)-dimensional Einstein–Hilbert action permits to integrate out all angular coordinates and eventually establishes an effective 2-dimensional model whose bulk action is precisely (2.15), with \( U(X) = -(D-3)/((D-2)X) \) and \( V(X) \propto X^{(D-4)/(D-2)} \). Curiously, in the limit \( D \to \infty \) the model derived from bosonic string theory is recovered (with 2-dimensional target space, see below) [67, 133, 134].

3. **Limiting case of Einstein–Hilbert in \( 2 + \epsilon \) dimensions.** Weinberg’s idea of asymptotic safety in gravity emerged from his consideration of gravity in \( 2 + \epsilon \) dimensions, in the limit of small \( \epsilon \) [135]. As we mentioned above, taking \( \epsilon \to 0 \) leads to trivial equations of motion. However, if simultaneously Newton’s constant scales to zero appropriately, then the limiting action can be non-trivial. In fact, Mann and Ross argued that the action obtained in this way is a 2-dimensional dilaton gravity action (2.15) with \( U(X) = \text{const.} \) and \( V(X) = 0 \) [136]. A more recent analysis confirms the result for \( U(X) \), but finds \( V(X) \propto e^{-2X} \) [137]. Such an action describes Liouville gravity, see [138, 139] for reviews.

4. **Higher power curvature theories.** Models that are non-linear in curvature and/or torsion are viable in two dimensions. In particular, the Katanaev–Volovich model describes 2-dimensional Poincaré gauge theory, i.e., a model with Lagrange density \( R^2 + T^2 \), where \( R \) is curvature and \( T \) torsion [140]. The Katanaev–Volovich model is classically equivalent to dilaton gravity (2.16) with \( U(X) = \text{const.} \) and \( V(X) \propto X^2 \), see [126, 141, 142]. Similarly, generic theories with non-linear Lagrangians in curvature and torsion are equivalent to generic dilaton gravity, provided the potentials \( U(X) \) and \( V(X) \) are chosen appropriately [143].

---

4 The derivation in [137] exploits a spherically symmetric ansatz in \( 2 + \epsilon \) dimensions, dualizes to a different action for which the limit \( \epsilon \to 0 \) is well-defined and dualizes back after taking the limit.
5. **Strings in two dimensions.** Conformal invariance of the sigma model action for the closed bosonic string,

\[ I^{(\sigma)} = \frac{1}{4\pi\alpha'} \int d^2 z \sqrt{-h} \left( g_{\mu\nu} h^{ij} \partial_i X^\mu \partial_j X^\nu + \alpha' \Phi \mathcal{R} \right) \]  

(2.17)

requires that the trace of the world-sheet energy-momentum tensor vanishes

\[ T_{i}^{i} \propto \beta \Phi \mathcal{R} + \beta_{\mu\nu} h^{ij} \partial_i X^\mu \partial_j X^\nu = 0 . \]  

(2.18)

The parameter \( \alpha' \) is the string tension, \( h_{ij} \) is the world-sheet metric, \( \mathcal{R} \) its Ricci scalar, \( X^\mu \) are the target space coordinates, and \( \Phi \) is the dilaton field. Thus, for consistency the \( \beta \)-functions appearing in (2.18) have to vanish [144].

\[ \beta_{\Phi} = -\frac{\alpha'}{4\pi^2} \left( \frac{26 - D}{12\alpha'} + (\partial \Phi)^2 - 4 \nabla^\mu \partial_\mu \Phi - \frac{1}{4} R \right) = 0 \]  

(2.19)

\[ \beta_{\mu\nu} = R_{\mu\nu} + 2 \nabla_\mu \partial_\nu \Phi = 0 \]  

(2.20)

Here \( D \) is the dimension of the target space, \( R_{\mu\nu} \) its Ricci tensor and \( \nabla_\mu \) the associated covariant derivative. The conditions of conformal invariance, \( \beta_{\Phi} = \beta_{\mu\nu} = 0 \), follow as equations of motion from a target space action, which for \( D = 2 \) turns out to be equivalent to dilaton gravity (2.15) with \( U(X) = -1/X \) and \( V(X) = 2\lambda^2 X \), upon identifying \( X = e^{-2\Phi} \). See [145–148] for some early literature on black holes in 2-dimensional string theory and [138, 149] for some reviews. The model by Callan, Giddings, Harvey and Strominger (CGHS) uses the same target space action as derived from string theory and adds matter fields to describe evaporating black holes [115]; the CGHS model engendered a lot of further research in 2-dimensional dilaton gravity with and without matter, see [67, 150, 151] for reviews.

Thermodynamics of 2-dimensional dilaton gravity models (2.15) was discussed assuming \( U(X) = 0 \) by Gegenberg et al. [152]. A comprehensive discussion of quasi-local thermodynamics for generic models (2.15) was provided using the Euclidean path integral approach more than a decade later [153].

Taken together, the body of results that these diverse two-dimensional models have in common suggests that certain features of black hole thermodynamics are universal. This is an important observation in its own right, independent of insights into the information paradox and other problems. In particular, with appropriate normalizations the ‘classical’ contribution to the entropy always takes the form

\[ S = 2\pi X_h \]  

(2.21)

where \( X_h \) is the value of the dilaton at the horizon. This result encapsulates inherently two-dimensional models, as well as the s-wave reduction down to two dimensions of the area law for higher dimensional theories. The robust nature of black hole
entropy was made apparent in the work of Wald, who gave a succinct geometric characterization of the entropy for any diffeomorphism-invariant theory of gravity [154, 155].

**Quasi-local thermodynamics and Hawking–Page phase transition.** A simple calculation shows that the Schwarzschild black hole has a negative specific heat, and therefore cannot be treated as an equilibrium thermodynamic system. York addressed the issue of negative specific heat by putting the black hole inside a cavity of some finite radius that provides a heat bath of fixed temperature [156]. For a sufficiently small cavity the specific heat is positive, leading to a well-defined canonical ensemble. It was shown later that some spacetimes, in particular asymptotically AdS spacetimes, naturally provide a covariant version of such a cavity. In all these examples the existence of a well-defined canonical ensemble means that interesting phase structures can be unraveled. Probably the most famous example is the Hawking-Page phase transition between “hot AdS”—anti-de Sitter space with periodic euclidean time \( \tau \sim \tau + T^{-1} \)—and an asymptotically AdS black hole [157]. For sufficiently small temperatures the minimum of the free energy is hot AdS, while at high temperatures the ensemble is dominated by the black hole.

**Gravity in three dimensions and a connection with conformal field theory.** During the same period there was a great deal of pioneering work in 3-dimensional gravity. Deser, Jackiw and Templeton constructed topologically massive gauge theories by adding a Chern–Simons term to the action [158–160]. In the case of gravity this leads to topologically massive gravity, a 3-dimensional theory of gravity that has a local (massive) gravitational degree of freedom. Its bulk action reads

\[
16\pi G I^\text{TMG} = \int d^3 x \sqrt{-g} (R - 2\Lambda) + \frac{1}{2\mu} \int d^3 x \varepsilon^{\mu\nu\lambda} \Gamma^\alpha_{\mu\beta} (\partial_\nu \Gamma^\beta_{\lambda\alpha} + \frac{2}{3} \Gamma^\beta_{\nu\gamma} \Gamma_{\gamma\lambda\alpha})
\]

(2.22)

Without the gravitational Chern–Simons term, \( \mu \to \infty \), Einstein gravity becomes locally trivial [161], but globally it can be non-trivial. In particular, in a seminal paper Brown and Henneaux found that the Hilbert space of any 3-dimensional theory of quantum gravity with AdS boundary conditions falls into representations of two copies of the Virasoro algebra, with central charges for Einstein gravity given by [162]

\[
c = \bar{c} = \frac{3\ell}{2G} \quad \text{where} \quad \Lambda = -\frac{1}{\ell^2}.
\]

(2.23)

This unexpected set of symmetries suggested that such theories might be amenable to an analysis using conformal field theory (CFT) techniques. The Brown–Henneaux results were an important precursor of the AdS/CFT correspondence found a decade later.

**Black holes in three dimensions.** Another crucial development was the discovery, by Bañados, Teitelboim and Zanelli (BTZ), of black hole solutions of 3-dimensional Einstein gravity with negative cosmological constant [163]. As discussed in [164], the BTZ black holes are locally AdS, but globally differ from AdS. In fact, they are certain orbifolds of AdS such that the ensuing solutions are locally AdS and remain
regular on and outside the event horizon. The line-element in ‘Boyer–Lindquist’ type coordinates ($\varphi \sim \varphi + 2\pi$),

\[
ds_{\text{BTZ}}^2 = -\left(\frac{r^2 - r_+^2}{r^2} - \frac{r^2 - r_-^2}{r^2}\right)\left(\frac{r^2 - r_+^2}{r^2} - \frac{r^2 - r_-^2}{r^2}\right) \frac{r^2 \ell^2}{(r^2)} \frac{dr^2}{dr^2} + r^2 \left(\frac{d\varphi + \frac{r_+ r_-}{\ell}}{d\varphi + \frac{r_+ r_-}{\ell}}\right)^2 \frac{dr^2}{dr^2}
\]

(2.24)

makes the similarity to rotating black holes in higher dimensions manifest: there is an ergosphere at $r = (r_+^2 + r_-^2)^{1/2}$, an outer horizon (with rotation) at $r = r_+$, an inner horizon at $r = r_-$, and a singularity behind the inner horizon. Moreover, there is a conserved mass, $M = (r_+^2 + r_-^2)/(8G\ell^2)$, and angular momentum, $J = r_+ r_-/(4G\ell)$. The presence of rotating black holes makes 3-dimensional AdS gravity a particularly interesting toy model to address classical and quantum aspects of black holes and their thermodynamical properties. In particular, the entropy is given by the Bekenstein–Hawking result (2.1)

\[
S_{\text{BTZ}} = \frac{2\pi r_+}{4G}.
\]

(2.25)

**Cardy formula.** The existence of the BTZ solution and the results of Brown and Henneaux led to the first attempt to explain black hole entropy by counting microscopic states. Since the Hilbert space of the theory is organized according to the symmetries of a two-dimensional CFT, one can carry out the state counting by exploiting a result of Cardy [165, 166]. Namely, given some assumptions there is a universal formula for the asymptotic density of states in a CFT. The log of the density of states leads to the Cardy formula for entropy

\[
S_{\text{Cardy}} = 2\pi \sqrt{\frac{c h}{6}} + 2\pi \sqrt{\frac{\bar{c} h}{6}},
\]

(2.26)

where $c, \bar{c}$ are the central charges and $h, \bar{h}$ are the Virasoro zero-mode charges. Evaluating the Cardy formula (2.26) for the Brown–Henneaux central charges (2.23) and the zero-mode charges $h = (\ell M + J)/2, \bar{h} = (\ell M - J)/2$ of the BTZ black hole (2.24) gives precisely the Bekenstein–Hawking entropy (2.25). This observation was the basis for the near horizon microstate counting pioneered by Strominger and Vafa a decade later [167, 168].

The explanation of (2.25) via microscopic state counting was a significant insight into the nature of black hole entropy. But it also left many important questions unanswered. In particular, the Cardy formula provides information about the asymptotic density of states but it gives no insight into the states themselves. An explanation of black hole entropy that proceeds from the identification of microscopic states would not be achieved until the following decade.

**Towards holography.** Counting black hole microstates through a CFT calculation is a remarkable manifestation of an idea that began to emerge at the end of the third
decade. In an essay dedicated to Abdus Salam, ‘t Hooft postulated that there is no information loss, i.e., the evolution describing collapse and quantum evaporation of a black hole should only incorporate processes that are not at odds with unitarity [169]. From this postulate and the observation that the Bekenstein–Hawking entropy (2.1) scales like the area, ‘t Hooft then argued that there could be an equivalent description of the system in terms of an ordinary quantum field theory in one dimension lower. A year later Susskind first coined the expression “holographic principle” and pointed out that string theory could be a candidate for a theory of quantum gravity realizing the holographic principle. But this part of the story already belongs to the next decade.

2.5 1993–2003

As described in the previous section, Cardy’s formula relates the Bekenstein-Hawking entropy of the BTZ black hole to the central charge of a two-dimensional CFT. This result foreshadows three major developments during the period 1993–2003: a complete accounting in string theory of microscopic states responsible for the entropy of certain black holes, the emergence of ‘t Hooft and Susskind’s holographic principle, and the development of the AdS/CFT correspondence as a fully-fledged example of holography.

Counting black hole microstates in string theory. String theory is a consistent theory of quantum gravity and is therefore a natural framework for investigating the microscopic origin of black hole entropy. As early as 1993, it was suggested that the density of states in perturbative string theory might be sufficient to explain the Bekenstein-Hawking entropy [170–172]. The main development during this period (and arguably one of the most significant accomplishments of string theory in any period) was Strominger and Vafa’s calculation of the density of states for certain supersymmetric black holes [173].

String theory contains both bosonic and fermionic degrees of freedom, with the bosonic sector including multiple $p$-form gauge fields under which black holes may be charged. In the case of supersymmetric black holes these charges completely characterize the horizon, which has an area that is independent of moduli like the string coupling or compactification volumes. The simplest such black holes involve either one or two charges, but such configurations possess either singular horizons or horizons with zero area. The Bekenstein-Hawking entropy is relevant for non-singular horizons with macroscopic area, which requires at least three charges. Strominger and Vafa considered these sorts of black holes in string theory compactified on the five dimensional product spaces $S^1 \times T^4$ and $S^1 \times K3$. Their construction involves $q_1$ D1-branes wrapping the circle, $q_5$ D5-branes wrapping all five compact dimensions, and massless strings stretched between the branes carrying $n$ units of momentum.

5 ‘t Hooft also provided as an example a realization of the holographic principle in terms of some cellular automaton model.
around the $S^1$. At weak coupling this system is described by a supersymmetric field theory on the worldvolume of the branes, and it is possible to enumerate the states with given charges. The resulting density of states is approximately
\[ \rho \approx \exp \left( 2\pi \sqrt{q_1 q_5 n} \right). \] (2.27)

As the gravitational (string) coupling is increased the picture changes, and at strong coupling the appropriate description of the system is a black hole. The horizon of this black hole has area
\[ A_H = 8\pi G \sqrt{q_1 q_5 n}, \] (2.28)

where each of the charges must be large to suppress various types of corrections. Although the descriptions at weak and strong coupling are radically different, the state counting is protected by supersymmetry. So even though the density of states (2.27) was derived at weak coupling, it still applies in the limit where the system is described by the black hole. To leading order the log of the density of states exactly reproduces the area law
\[ S = \log \rho = 2\pi \sqrt{q_1 q_5 n} = \frac{A_H}{4G}. \] (2.29)

Thus, this result of Strominger and Vafa provides the first derivation of the Bekenstein–Hawking entropy that identifies and counts a specific set of microscopic states associated with the parameters describing the macroscopic black hole. Similar calculations have been carried out for supersymmetric black holes in four dimensions [174], near-extremal black holes [175, 176], and even certain extremal black holes with broken supersymmetry [177]. Comprehensive reviews can be found in [178, 179].

Despite the success of this program, there is still no explicit construction of the microstates of non-supersymmetric, non-extremal black holes like the Schwarzschild or Kerr solutions (though, in the latter case progress has been made for the extremal solution [180]). It is also important to point out that while the counting of states is protected by supersymmetry, the states in the strong coupling regime bear no resemblance to the states at weak coupling. In this sense, it is not clear what constitutes the “states of the black hole”. Indeed, given a generic state in the weakly coupled regime it is not clear what happens as the coupling is increased. It is possible (and with hindsight also plausible) that the states in the strongly coupled regime are free of horizons. This idea has motivated a tremendous amount of work—the so-called microstate and fuzzball programs—which will be discussed in the next section.

**Holographic principle.** Around the same time that a stringy origin for the black hole density of states was first being considered, ’t Hooft put forth a radical suggestion: that gravitational physics in $3 + 1$ dimensions must effectively become $2 + 1$ dimensional at Planckian scales [169]. Susskind, building off his own work on the role of string theory in explaining the Bekenstein–Hawking entropy, explored the consequences of this idea and dubbed it the “holographic principle” [181]. This principle
is often regarded as synonymous with the Bekenstein–Hawking area law for black hole entropy, but it is in fact a much deeper statement about locality, unitarity, and the nature of quantum gravitational physics.

In its earliest form, the holographic principle was interpreted as a bound on the number of degrees of freedom needed to describe physics in a spatial region. Quantum field theory suggests that any such region contains an infinite number of degrees of freedom associated with the infinite number of harmonic oscillator states possible at each of the infinite number of points in the region. Including gravity changes this counting, since exciting too many of these states would provide enough energy to form a black hole. A better estimate would ‘coarse grain’ space on lengths of order the Planck scale and, at the very least, place an upper limit on the energy contained in any Planck volume to avoid creating a black hole. With these restrictions the number of degrees of freedom scales like the volume $V$ of the region. But this must be a gross over-counting, since black holes could still form on larger scales even if the energy bound on each Planck volume was not saturated. And since the largest black hole that ‘fits’ in the region has an entropy given by $A/4$, it must be that the number of accessible degrees of freedom in a region scales like the area bounding the region rather than its volume.

The conclusion described above forces a choice between locality and unitarity. If all the degrees of freedom predicted by local physics were available in a region of volume $V$, then it would not be possible to accommodate all possible states of the system with the dramatically reduced number of states after gravitational collapse. To preserve unitarity, it must be that physics in any region bounded by a surface of area $A$ is described by no more than $A/4$ degrees of freedom, even in the absence of a black hole.

This early form of the holographic principle depends crucially on the idea that the entropy in a spatial region $V$ is limited by the area of the surface $B = \partial V$ bounding the region

$$S[V] \leq \frac{c^3}{4G\hbar} A(B) . \quad (2.30)$$

But it was soon realized that this spacelike form of the entropy bound can fail [182, 183], leading researchers to attempt a reformulation of the bound in terms of light cones. This program culminated with Bousso’s Covariant Entropy Conjecture [184], a covariant generalization of the original bound which replaces the spacelike region $V$ with a null hypersurface. Specifically, given some surface $B$ with area $A(B)$ the light sheet $L(B)$ is the null hypersurface generated by following light rays from $B$ until they begin to expand. The entropy on any light sheet of a surface is then bounded according to

$$S[L(B)] \leq \frac{c^3}{4G\hbar} A(B) . \quad (2.31)$$

A comprehensive review of the Covariant Entropy Conjecture and the holographic principle in general is given in [185].
Like other entropy bounds, there is no formal derivation of (2.31). Rather, it is a conjecture for which there is strong circumstantial evidence and a lack of counterexamples. Since any derivation of this result would require a complete theory of quantum gravity, it is hoped that the holographic principle will instead provide some guidance as to what such a theory might be. It is tempting, given the form of the bounds (2.30) and (2.31), to assume that the physics interior to a region is somehow encoded on its boundary. The holographic principle offers little direct insight as to whether this is the case, or how it might be accomplished\(^6\). Nevertheless, this assumption, combined with calculations inspired by the work of Strominger and Vafa, leads to a fully realized form of the holographic principle in Maldacena’s AdS/CFT correspondence.

**AdS/CFT correspondence.** The work of Strominger–Vafa showed how the entropy of certain supersymmetric black holes may be understood via a calculation in a field theory on the world volume of a D-brane bound state. The entropy is not the only quantity that can be explained this way. For instance, absorption cross sections calculated using both the gravity and field theory descriptions are found to agree. This observation inspired similar comparisons for a stack of D3-branes in type IIB string theory \([186–188]\). The agreement between the gravity and field theory calculations for the D3-brane system gives the first pieces of evidence for the AdS/CFT correspondence.

Given a stack of \(N\) parallel D3-branes, low energy excitations on the worldvolume are described by a four-dimensional \(U(N)\) gauge theory with \(\mathcal{N} = 4\) supersymmetry \([189]\) and a coupling constant related to the string coupling by \(g_{YM}^2 \sim g_s\). For \(N\) large and \(g_{YM}^2 N \ll 1\) the theory is well-described by perturbation theory with non-planar diagrams suppressed by factors of \(1/N\). On the other hand, the near-horizon geometry of the stack of branes looks like a product space of the form \(\text{AdS}_5 \times S^5\), with both factors having a radius of curvature \(\ell\) that satisfies

\[
\ell^4 = 4\pi g_s N (\alpha')^2. \tag{2.32}
\]

The description of the system in terms of gravitational physics requires curvatures to be much smaller than the string scale, \(\ell \gg \sqrt{\alpha'}\), which implies \(g_s N \gg 1\). In other words, the gravitational description can be trusted precisely when the worldvolume field theory is strongly coupled. Maldacena conjectured that these descriptions are in fact the same; two sides of a strong-weak coupling duality \([190]\). In this picture the conformal symmetries of the field theory are realized by the \(SO(4,2)\) isometries of \(\text{AdS}_5\), while the R-symmetries are encoded in the \(SO(6)\) symmetries of the \(S^5\).

The strongest form of Maldacena’s conjecture asserts that type IIB string theory with \(\text{AdS}_5 \times S^5\) boundary conditions is completely equivalent to four-dimensional Super Yang-Mills for all values of the parameters \(g_s\) and \(N\). This is the most tantalizing and least tested form of the correspondence. When \(N \to \infty\) at fixed \(g_{YM}^2 N\) the duality relates classical string theory to Super Yang-Mills with finite coupling, and

\[^6\text{Such an encoding results in an entropy that scales like the area, which suggests a local and non-gravitational description on the boundary.}\]
many consequences of this form of the conjecture have been tested using unexpected
integrability properties of the planar sector of SYM [191]. The weakest and most
thoroughly examined form of the conjecture follows from letting $g_{YM}^2 N \to \infty$. In
that case the gravitational side of the duality reduces simply to type IIB supergravity
on $\text{AdS}_5 \times S^5$, which is equivalent to the strong-coupling limit of SYM. All forms of
the duality are manifestly holographic, in the sense that the gravitational physics of
a $d + 1$ dimensional asymptotically AdS spacetime is encoded in a local field theory
on the spacetime’s $d$ dimensional conformal boundary.

Maldacena’s original conjecture, which includes a number of other brane config-
urations with low-energy descriptions in terms of various supergravities, has been
extended, deformed, and modified in various ways. It has primarily been used to
extract useful statements about strongly coupled gauge theories. For instance, cor-
relation functions of operators in the gauge theory can be calculated from the string
theory partition function, which in the standard (weak) form of the correspondence
is dominated by contributions from saddle points of the supergravity action. The
on-shell action can be expressed as a functional of ‘boundary data’ $\phi_0$ for the fields
$\phi$ that play the role of sources $J$ for operators $O$ in the dual field theory

$$Z_{\text{sugra}}[\phi_0 = 0 | \partial \text{AdS}] = Z_{\text{CFT}}[\phi_0 = J] \sim \langle \exp \left( \int O \phi_0 \right) \rangle_{\text{CFT}}. \quad (2.33)$$

The full impact of AdS/CFT on the study of strongly coupled gauge theories is
beyond both the purpose and scope of this review. But the duality does offer several
useful insights into black hole thermodynamics, which we will focus on for the rest
of this section.

Not long after the AdS/CFT correspondence was first proposed, Witten showed
how the thermodynamics of an AdS black hole can be understood in terms of the
(large $N$) thermodynamics of the dual gauge theory [192]. In particular, the usual
Hawking–Page transition from AdS-Schwarzschild to “hot AdS” corresponds to a
confining/deconfining phase transition in the dual field theory. This can be seen
from the free energy of the two bulk configurations, which when expressed in terms
of field theory quantities scales as $F \sim O(1)$ and $F \sim O(N^2)$, respectively, in the
confined and deconfined phases.

The AdS/CFT correspondence also illuminates calculations of the entropy of the
BTZ black hole, raising Brown and Henneaux’s result [162] from an analogy to an
actual counting of states in a dual CFT [167, 193]. This is especially important for a
number of black holes that arise in string theory, which typically have near-horizon
geometries of the form $\text{BTZ} \times Y$ for some space (or product of spaces) $Y$. The entropy
of these black holes can then be explained via a similar state counting without having
to work out the full details in the worldvolume theory. For a review, see [178, 179].

Perhaps the most important consequence of AdS/CFT for black hole thermo-
dynamics is the idea that a gravitational theory, which presumably includes black

---

7 The dual field theory at finite temperature is defined on $S^3 \times S^1$ and therefore has compact volume.
Nevertheless, a phase transition is possible because the theory is considered in the large $N$ limit.
holes, is equivalent to a theory that is unitary. There are many ways to interpret such a statement in the context of the information paradox. Since the duality applies to the dynamics of both theories it is tempting to ‘resolve’ the paradox by pointing out that any process on the gravity side—including the formation and eventual evaporation of a black hole—is encoded in unitary physics on the field theory side. But this is far from a complete argument. In particular, the unitary evaporation of an AdS-Schwarzschild black hole still forces one to either abandon local Hamiltonian evolution (in the bulk) in a setting where it is expected to be a good description, accept the formation of some sort of macroscopic remnant that remains entangled with the Hawking radiation, or else revisit assumptions about the formation of black holes in string theory [194]. One possible resolution is that the weakly coupled D-brane states that are counted in, for example, the Strominger–Vafa calculation do not form horizons as the gravitational coupling is increased. Instead, such states possess significant structure on horizon scales, and the traditional black hole is viewed as a coarse-grained description of the actual states. This possibility, which was mentioned earlier, is the basis for the microstate and fuzzball programs described in the next section.

The AdS/CFT correspondence is, at present, the most fully realized implementation of the holographic principle. It therefore owes its existence, at least in part, to the comparatively humble idea that the entropy of a black hole scales like the horizon area (2.1). In turn, AdS/CFT has inspired a number of generalizations, extensions, and applications which may be considered descendants of black hole thermodynamics. Some early examples during the period 1993–2003 include duals of confining field theories with \( N = 1 \) supersymmetry [195], the dS/CFT correspondence relating quantum gravity on de Sitter space to a Euclidean CFT [196, 197], proposed duals of \( O(N) \) vector models in terms of higher spin gauge theories [198], and even applications of gauge/gravity duality techniques to calculations in inflationary cosmology [199–201].

The topics discussed in this section represent major achievements during the period 1993–2003, but they were certainly not the only interesting developments during that time. For instance, in 1995 Jacobson was able to extract, under certain assumptions, the Einstein equations from horizon thermodynamics [202]. This result inspired a fair amount of subsequent work, especially in recent years [203, 204].

2.6 2003–2013

The previous decade saw great progress in microscopic state counting, and the emergence of holography as an important and perhaps fundamental property of quantum gravity. In recent years there has been a focus on applications and generalizations of AdS/CFT, efforts to identify the gravitational states associated with a black hole, and attempts to comprehensively resolve the information paradox. Some new problems have arisen, but developments that touch on two or more of these issues suggest
a convergence towards a deeper understanding of quantum gravity and black hole thermodynamics.

**Tests and applications of AdS/CFT.** Early tests of AdS/CFT spawned a number of further checks that probed different regimes of the correspondence. For instance, methods known from integrable systems, such as the thermodynamic Bethe Ansatz, allowed to check aspects of AdS/CFT beyond perturbation theory (in particular, for arbitrary values of the 't Hooft coupling constant $\lambda$); see [191] for a review. As the correspondence matured a number of new applications were uncovered. An emblematic example is the prediction of the ratio of shear viscosity $\eta$ to entropy density $s$ in the infinite coupling limit [205, 206].

$$\frac{\eta}{s} = \frac{\hbar}{4\pi k_B}$$

In relativistic heavy ion collisions the same order of magnitude was observed for $\eta/s$ (see [207]), which inspired both phenomenologists and theoreticians to apply AdS/CFT methods to the description of relativistic plasmas, see e.g. [208–211] for reviews. The key feature of the $\eta/s$ story is that a complicated calculation on the field theory side—determining the shear viscosity for a strongly coupled plasma—is mapped to a problem on the gravity side that is suitable for a bright PhD student. Indeed, Damour provided a comparable calculation in his PhD thesis already in 1979 [212].

**Gauge/gravity correspondences.** The past decade has seen numerous further attempts to phenomenologically apply ideas from the AdS/CFT correspondence to more general settings. These ‘gauge/gravity’ correspondences began with deformations of AdS/CFT, but were soon extended to conjectured dualities between theories that bear little resemblance to asymptotically AdS gravity or $\mathcal{N} = 4$ Super-Yang–Mills. As above, the idea is to map complicated (strong coupling) problems on the gauge theory side to fairly simple problems on the gravity side. Examples include condensed matter applications such as cold atoms [213–215], Lifshitz fixed points with non-relativistic scaling symmetries [216], superfluids/superconductors [217–219], non-Fermi liquids/strange metals [220–223] and the gravity/fluid correspondence [224–228] (based on the membrane paradigm [229]). Some applications of proposed gauge/gravity dualities to condensed matter systems are reviewed in [230–235].

**Limits of holography.** We discuss now in a bit more detail some extensions of the AdS/CFT correspondence that are more in line with the main topic of our review. An interesting theoretically motivated question to ask is, how general is holography? Originally, the holographic principle was motivated by avoidance of information loss and preservation of unitarity, but the way the AdS/CFT correspondence works makes it plausible that it could also apply to systems that are non-unitary. Moreover, if the holographic principle is a true statement about Nature then it should be realized in settings other than AdS, such as asymptotically flat or accelerating Friedmann–Lemaître–Robertson–Walker spacetimes.
Finally, it is interesting to ask whether there are theories apart from string theory that permit a holographic description. A conclusive answer to this question would be an important achievement. If affirmative, then such theories might provide novel playgrounds for theoretical considerations about holography as well as new applications along the lines of AdS/CFT. If negative, we would have established a direct link between holography and string theory, i.e., holography would necessarily imply string theory.

**Non-unitary holography.** Partly for simplicity and partly because there were many developments in the past decade, we restrict ourselves mostly to 3-dimensional theories of gravity in order to address the issues raised in the previous paragraph. Let us start with the question to what extent holography could apply to non-unitary theories. This question is somewhat delicate, because non-unitarity is often associated with some sickness of the theory. However, there are also systems that exhibit non-unitarity in a ‘controlled’ way. This includes, for instance, open quantum systems and systems with quenched disorder. In a story with several interesting twists, it appears that TMG (2.22) at the critical point $\mu \ell = 1$ corresponds to a log CFT, as suggested first in [236]. Log CFTs are specific non-unitary CFTs where two or more operators have degenerate scaling dimensions and the Hamiltonian acquires a Jordan block structure [237–239]. They are used, for example, in the description of systems with quenched disorder. A key element on the gravity side is the emergence of log modes [236]

$$\psi_{a\beta}^{\log} = \lim_{\varepsilon \to 0} \frac{\psi_{a\beta}^{M} - \psi_{a\beta}^{L}}{\varepsilon} = -2 (i t + \ln \cosh \rho) \psi_{a\beta}^{L}$$ (2.35)

as linearized perturbation on the AdS background. The middle equation indicates the degeneration of the massive graviton modes $\psi^{M}$ with the ‘left-moving boundary graviton’ modes $\psi^{L}$ (specific Einstein-gravity modes at linearized level). The latter are eigenstates of the Hamiltonian, $H \psi^{L} = i \partial_t \psi^{L} = h \psi^{L}$, while the former are not: $H \psi^{\log} = i \partial_t \psi^{\log} = h \psi^{\log} + 2 \psi^{L}$. These two equations make manifest the Jordan block structure of the Hamiltonian $H$ when acting on the pair $\psi^{\log}, \psi^{L}$. See [240] for a full account of various checks, generalizations and possible applications of the AdS/log CFT correspondence. Thus, it seems that it is possible to extend the holographic principle to theories that exhibit non-unitarity in a controlled way.

**Flat space holography.** There was some progress on extracting features of the flat space S-matrix from AdS/CFT correlators, see e.g. [241–245], but it is still fair to say that efforts at flat-space holography have not met with a great deal of success in dimension four and above. In three dimensions one can essentially repeat the Brown–Henneaux construction, which was done by Barnich and Compere [246]. The asymptotic symmetry algebra was found to be the Bondi–van der Burg–Metzner–Sachs (BMS) algebra [247, 248] in three dimensions, which arises also as the ultra-relativistic contraction (or large AdS radius limit) of the two-dimensional conformal algebra. These algebras are also known as Galilean conformal algebras [249, 250], which led to the notion of a ‘BMS/GCA correspondence’ [251]. A specific proposal
for a flat space/CFT correspondence is flat space chiral gravity (TMG (2.22) in
the limit $\ell \to \infty$ and $G \to \infty$, with $\mu G$ kept finite), which is conjectured to be
dual to a chiral half of a CFT [252]. In particular, for central charge $c = 24$ the
conjectured CFT is a chiral half of the monster CFT (proposed originally by Witten
in the context of Einstein gravity [253]), exactly like in the chiral gravity proposal
by Li et al. [254]. Its partition function is given by the $J$-function [253, 255] and due
to chirality depends solely on the ‘left-moving’ modular parameter $q$.

$$Z(q) = J(q) = \frac{1}{q} + 196884 q + O(q^2)$$

(2.36)

The number 196884 is interpreted as one Virasoro descendant of the vacuum plus
196883 primary fields corresponding to flat space cosmology horizon microstates
(see below). The flat space chiral gravity quantum entropy $S = \ln 196883 \approx 12.2$
diffs only by about 3 % from the semi-classical Bekenstein–Hawking result $S_{BH} =
4\pi \approx 12.6$ (in suitable units). For quantum gravity applications the (flat space)
chiral gravity situation seems optimal: there are quantum corrections that are not
completely negligible (of the order of a few percent), but the theory is not “ultra-
quantum” so that geometric notions associated with the semi-classical limit, like
black hole horizons, can still be discussed meaningfully. Flat space cosmologies
[256, 257] are the flat space analog of BTZ black holes (2.24) and permit a microstate
counting similar to AdS (2.26), see [258, 259]. They are subject to a Hawking–Page
like phase transition [260] so that at least in three dimensions cosmic evolution can
be generated by heating (and gently stirring) flat space. For further aspects of flat
space holography see e.g. [261–269].

**Higher spin holography.** Remarkably, AdS spacetimes permit interacting massless
particles with spin greater than two [270–273]. These ‘higher spin’ theories could be
relevant for the holographic description of certain sectors of large $N$ gauge theories.
In particular, Klebanov and Polyakov proposed that a particular Vasiliev-type higher
spin theory on $AdS_4$ might be exactly dual to the $O(N)$ vector model (at large $N$)
in three dimensions. This conjecture triggered an intensive study of the subject with
impressive achievements [274–277]. An interesting technical aspect of higher spin
holography is that it provides a weak/weak duality and therefore allows to test holog-
ography with high precision (the other side of the coin is that higher spin holography
is of less practical use than AdS/CFT, since strong/weak dualities can map hard
calculations to simple ones). Coming back to three bulk dimensions, Henneaux and
Rey (and independently Campoleoni, Fredenhagen, Pfenninger and Theisen) gen-
eralized the Brown–Henneaux analysis to higher spin theories with AdS boundary
conditions [281, 282], and a few months later Gaberdiel and Gopakumar proposed a
correspondence between Vasiliev-type higher spin gravity and minimal model CFTs
[278–280]. Some selected papers and reviews are [283–292]. Recently, the topics
of flat space and higher spin holography were combined [293, 294], in the spirit
of non-AdS holography for three dimensional higher spin gravity [295]. The main
observation is that unlike the spin-2 case, higher spin theories allow for many different
backgrounds, including Lobachevsky, Lifshitz, Schrödinger and warped AdS besides more common backgrounds such as AdS or flat space, without the addition of matter fields.

**Holographic entanglement entropy.** Entanglement entropy is an entanglement measure for bipartite pure states $|\Psi\rangle$ and is defined as the von Neumann entropy

$$S_A = -\text{tr}\rho_A \ln \rho_A = -\lim_{n \to 1} \frac{d}{dn} \text{tr}\rho^n_A$$

(2.37)

associated with the reduced density matrix $\rho_A = \text{tr}_B |\Psi\rangle\langle\Psi|$, where the total system is divided into two subsystems $A$ and $B$, see e.g. [296]. For the present context one can think of $A(B)$ as the exterior (interior) of a black hole. Then $S_A$ can be thought of as the entropy for an observer who has access only to the black hole exterior. The fact that entanglement entropy obeys an area law led to the suggestion [297–299] that Bekenstein–Hawking entropy could be interpreted as entanglement entropy (see [300] for a review).

Entanglement entropy has found many applications in quantum systems, see e.g. [301–304], but is not easy to calculate in interacting quantum field theories in dimension greater than two [sometimes the so-called replica trick can be used, which exploits the second equality in (2.37)]. The holographic entanglement entropy proposal by Ryu and Takayanagi [305, 306] applies holographic ideas to map the difficult calculation of entanglement entropy on the field theory side to an elementary calculation of minimal surfaces on the gravity side. This proposal has passed several tests by successfully reproducing the entanglement entropy in well-understood cases, see [307] for a review. Taking the proposal for granted it can then be applied to situations in which no other method exists (currently) to determine entanglement entropy. Thus, holographic entanglement entropy, which provides a link between black hole thermodynamics and quantum information, is another example of the utility of weak/strong dualities like AdS/CFT.

**Geometry of black hole thermodynamics and cosmological constant as state parameter.** Over the last decade most of the work inspired by black hole thermodynamics focused on holography, AdS/CFT, and related issues. But there were also some interesting purely thermodynamical developments. For example, the geometry of black hole thermodynamics was investigated in numerous papers, see for instance [310–316]. The basic idea goes back to Ruppeiner [317–319], namely to associate the Hessian of the entropy (with respect to some state space variables $x^i$) with a metric,

$$g_{ij} = -\partial_i \partial_j S(x^k),$$

whose geometric properties are related to the thermodynamics of the system.

Another example is the recent revival of the idea to treat the cosmological constant as a state parameter. This concept goes back to the germinal work of [320, 321]. In order to treat the cosmological constant as a state variable, $\Lambda$ is introduced as a state variable.

---

8 We mention in the conclusions that this link is likely to grow stronger in the future. Besides the numerous recent papers on holographic entanglement entropy, some selected papers that also provide such links are [308, 309] and references therein.
constant of integration by coupling a four form field strength to gravity. The value of the cosmological constant can change by spontaneous nucleation of membranes that act as sources for the four form \[322–324\], or by thermal decay together with the creation of a black hole \[325\]. These results motivate the study of black hole thermodynamics in AdS, in a phase space extended by $\Lambda$ and its conjugate variable $\Theta$, the negative of which turns out to be a suitable “thermodynamic” definition for the volume of a black hole \[326–333\]. See \[334\] for a review.

**Kerr/CFT.** The counting of black hole microstates pioneered by Strominger and Vafa in the previous decade initially was restricted to simple but astrophysically irrelevant black holes. In the decade discussed in this section a similar counting was applied to Kerr black holes, which established the ‘Kerr/CFT’ correspondence, see e.g. \[335–345\]. Particularly the early papers were based on the near horizon extremal Kerr (NHEK) metric constructed by Bardeen and Horowitz \[346\].

\[
ds^2_{\text{NHEK}} = M^2(1 + \cos^2 \theta) \left( \frac{-d\hat{t}^2 + d\hat{r}^2}{\hat{r}^2} + \frac{4 \sin^2 \theta}{(1 + \cos^2 \theta)^2} \left( d\hat{\phi} + \frac{d\hat{t}}{\hat{r}} \right)^2 + d\theta^2 \right) \quad (2.38)
\]

The line-element \(2.38\) is obtained from the Kerr geometry \(2.2\) by rescaling $\hat{t} = \frac{\lambda t}{2M}$, $\hat{r} = \frac{\lambda r - M}{2M}$, $\hat{\phi} = \phi - \frac{t}{2M}$ and taking the limit $\lambda \to 0$ while keeping $\hat{t}$, $\hat{r}$, $\hat{\phi}$, $\theta$ fixed. The entropy counted by CFT methods then matches the Bekenstein–Hawking result \(2.1\).

\[
S_{\text{CFT}} = \frac{2\pi J}{\hbar} = \frac{A_h}{4G} = S_{\text{BH}} \quad (2.39)
\]

While astrophysical black holes are never exactly extremal (the Thorne bound on the dimensionless Kerr parameter is $a < 0.998 \ [347]$), some of them come very close to this bound. A possible example is GRS1915+105 whose dimensionless Kerr parameter appears to exceed $a \gtrsim 0.98 \ [348, 349]$ (however, see \[350\]).

**Fuzzballs.** The various successes of counting black hole microstates all failed to answer an important question: what do the corresponding microstate geometries look like? The fuzzball proposal \[351\] addresses this question in the context of string theory, stating that there should be $\mathcal{O}(e^S)$ horizonless and regular solutions that asymptote to the geometry of a given black hole, but differ from this geometry at the scale of the horizon, see also \[352, 353\], and \[354–357\] for reviews. The fuzzball proposal is motivated by the AdS/CFT correspondence as follows: for every state in the CFT counted by the Cardy formula there is a corresponding regular asymptotically AdS geometry. Each of these geometries encodes the vacuum expectation values of gauge invariant operators in that state through the standard AdS/CFT dictionary. These solutions can be stringy in the interior, though large classes of solutions have been identified that are well-described by the supergravity approximation.

One of the main achievements claimed by proponents of the fuzzball program is a resolution of the information paradox. This can be traced back to a key property of the proposal, which is that quantum gravity effects in string theory can take place on scales much larger than the Planck scale due to ‘fractionation’ \[194\]. This results in
significant modifications of Hawking radiation at wavelengths of order $GM$, allowing information to escape the ‘hole’ and be recovered (in principle) by external observers. On the other hand, if the system is probed with some object of sufficiently high energy, $E \gg T$ (where $T$ is the Hawking temperature), then collective modes of the fuzzball are excited, which is well-approximated by a description in terms of an ensemble average over all fuzzballs. The latter reproduces the black hole geometry, so that the dynamics of sufficiently energetic objects over short timescales (like an astronaut falling into a black hole) are essentially the same as one would expect in a classical black hole geometry.

**Firewalls.** The information loss problem has resurfaced in the past few years through an ingenious gedankenexperiment set up by Almheiri, Marolf, Polchinski and Sully (AMPS) that highlighted their difficulty of reconciling black hole complementarity with the equivalence principle [358]. AMPS and several other authors argued that a possible resolution of this incompatibility results in an infalling observer encountering a ‘firewall’ close to the horizon, see [359, 360] and references therein. The AMPS gedankenexperiment has engendered a lot of discussion and is an excellent demonstration that, at least collectively, the days of confusion regarding black hole thermodynamics and information loss are not over yet. There is of course a simple resolution of the apparent firewall paradox, but the margin is too small to include it here.

### 2.7 Conclusions and Future

**Log corrections to entropy.** We started our journey through the past 50 years of black hole thermodynamics with the Bekenstein–Hawking relation (2.1) and the statement that black hole thermodynamics provides non-trivial consistency checks for quantum gravity. We will end our review on a similar note, by going one step further than Bekenstein and Hawking. Namely, in the semi-classical approximation the area law obtains quantum corrections, which can be organized in an expansion in terms of $1/S_{\text{BH}}$ (the same kind of correction is obtained from subleading contributions to the Cardy formula [361]).

$$S = S_{\text{BH}} + \gamma_1 \ln S_{\text{BH}} + \gamma_2 + \mathcal{O}(1/S_{\text{BH}})$$

(2.40)

While the subleading terms and the $\mathcal{O}(1)$ term depend on the specific quantum gravity theory, the leading and first subleading term depend only on the classical limit of that theory and the validity of the semi-classical approximation. In other words, any theory of quantum gravity that is supposed to be equivalent to Einstein gravity in its semi-classical limit must not only reproduce the Bekenstein–Hawking law (2.1), but also the same result for the numerical coefficient $\gamma_1$ in front of the logarithmic correction term as obtained from perturbative (1-loop) quantization of Einstein gravity (with a given set of matter degrees of freedom—the precise coef-
ficient depends on the specific matter content). A recent summary of logarithmic corrections to Schwarzschild and other non-extremal black holes was provided by Sen [362]. He found that string theory calculations, whenever available, agree precisely with the semi-classical result. Interestingly, the simplest of all black holes, the Schwarzschild black hole, still presents a challenge: currently, string theory does not provide a prediction for $\gamma_1$ of the Schwarzschild black hole. 9

**Future developments.** Predictions of future developments often serve as a source of amusement for future generations [365], but we will venture one as our closing statement. While a lot of our current understanding of black hole thermodynamics and quantum gravity was achieved through consistently applying Feynman’s dictum “everything is particle”—most prominently epitomized by the Hawking effect—we predict that most of our future understanding will be achieved through consistently applying Wheeler’s dictum “everything is information” [366], like in the recent slogan “ER = EPR” [367] that emerged from the firewall discussions.

**Acknowledgments** DG and RM thank their respective collaborators for numerous discussions on black hole thermodynamics in the past 15 years. DG and JS were supported by the START project Y 435-N16 of the Austrian Science Fund (FWF) and the FWF projects I 952-N16 and I 1030-N27.

**References**

13. Doroshkevich, A., Zel’dovich, Y., Novikov I.: Gravitational collapse of nonsymmetric and rotating masses. JETP **49** (1965)

---

9 Loop quantum gravity does provide such a prediction [363, 364], and it disagrees with the semi-classical result.


365. Watson, T.: I think there is a world market for maybe five computers. (president of IBM), (1943)

366. Wheeler, J.: Everything is information or “It from bit” was a slogan that replaced Wheeler’s earlier belief everything is geometry. Indeed, taking geometry too seriously can be misleading in the context of quantum gravity. (1990)

Quantum Aspects of Black Holes
Calmet, X. (Ed.)
2015, xi, 322 p. 57 illus., 23 illus. in color., Hardcover
ISBN: 978-3-319-10851-3