Imperial College London

Black Holes and the Unitarity of Hawking Radiation

MSc in Quantum Fields and Fundamental Forces
Dissertation

Andrés Bañares Hernández[1]

Supervisor: Prof. Toby Wiseman

Theoretical Physics Group
Blackett Laboratory
Imperial College

London, SW7 2AZ, U.K.

25th September, 2021

[1] andres.banares-hernandez20@imperial.ac.uk

To my parents, whom I am deeply grateful to

Abstract

We review the progress in the understanding of how the entropy of black holes and their associated radiation is computed, and whether this conflicts with unitarity or not. In particular, we shall start with Hawking and Bekenstein's calculations of evaporating black holes and information in the 70s. We will then move to more recent progress derived from generalisations of this in the context of holography, leading to the formulation of a prescription for computing the entropy of quantum gravitational systems that correctly reproduces Page curves and their associated unitary evolution.

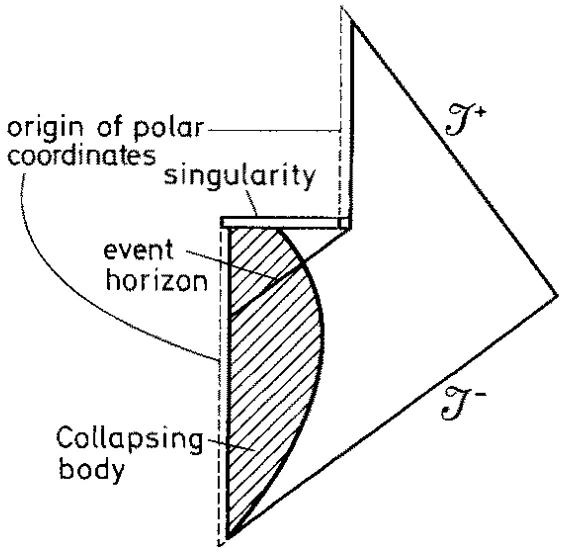


Image credit: [1]

Figure 0.0.1: Hawking's depiction of a Penrose diagram for an evaporating black hole used in his original calculation.

Contents

1 Introduction **1**
1.1 Hawking Radiation . . . 1
1.2 Unitarity and the Page Curve . . . 4
1.3 Bekenstein-Hawking Entropy and the Holographic Principle . . . 5
1.4 Fine-grained Entropy of Quantum Systems . . . 9

2 The AdS/CFT Correspondence **13**
2.1 Anti-de Sitter Space . . . 13
2.2 Conformal Field Theory . . . 18
2.3 Statement of the AdS/CFT Correspondence . . . 21

3 A General Expression for the Gravitational Entropy **23**
3.1 Counting Degrees of Freedom in AdS/CFT . . . 23
3.2 The Ryu-Takayanagi Formula . . . 26
3.3 HRT Proposal and General Covariance of the Entropy . . . 31
3.4 FLM/EW Proposals and Quantum Extremal Islands . . . 35

4 Recovering the Page Curve **40**
4.1 A Prescription for Evaporating Black Holes . . . 40
4.2 The Entropy of a Black Hole . . . 42
4.3 The Entropy of Radiation . . . 47
4.4 The Entanglement Wedge . . . 50

5 Two-dimensional JT Gravity **52**

6 Further Comments **59**

7 Acknowledgements **59**

References **59**

1 Introduction

1.1 Hawking Radiation

In Hawking's breakthrough paper [1], it was shown how by considering quantum field theory on a static Schwarzchild background, one was led to the natural conclusion that black holes are inherently thermal objects: The QFT vacuum is unstable and entangled modes of radiation are created near the horizon, which effectively means that black holes should evaporate and have a temperature.

In particular, one can can think of the simplified description of a particle / anti-particle pair being created, one ingoing and another outgoing. The ingoing one gets absorbed by the black hole and reduces its energy, and thus its mass and its size, which is characterized by the Schwarzschild radius of $r_s = 2G_N M$. The outgoing particle, on the other hand, leaves the black hole and can be detected by an observer in the form of radiation.

This novel idea led to a profound and once unsuspected connection between the physics of gravitational systems and thermodynamics, which is not manifest in classical general relativity. It was perhaps what may be considered as the first successful attempt to reconcile some aspects of the two great pillars of modern physics that were developed in the 20th century: the gravitational description of relativity and the microscopic world described by quantum mechanics. Indeed, while a full theory of quantum gravity remains a highly challenging problem to this day, Hawking's contribution remains highly relevant, which stands as a testament to the depth of Hawking's insights.

Hawking's remarkable formula for the temperature of a black hole, in a very succinct way summarizes the great unity of the concepts it relates: gravity, thermodynamics and quantum mechanics. It is given by

$$T_H = \frac{\hbar c^3}{8\pi G_N M k_B}, \tag{1.1.1}$$

while we will typically work with natural units, we have purposefully reinstated the universal constants it relates to make this unity manifest.

If some historians today with elementary knowledge of physics were to walk into Westminster Abbey in London, they would see this formula engraved in Hawking's tombstone. What's more, a few steps away in the same place, they would see Newton's tombstone, whose constant also appears in the formula. If they later went to a graveyard in Vienna, they would see Boltzmann's formula, whose constant too appears in Hawking's formula. It therefore seems likely that they would come to the conclusion that Hawking's formula was quite unique, even if they had no idea of what it described, which is also quite remarkable by itself.

Following [2], there is a relatively simple way of arriving at this formula, which also lends some insights on its nature. To do this, consider the usual Schwarzschild metric for a spacetime outside a black hole, i.e.

$$ds^2 = -\left(1 - \frac{r_s}{r}\right) dt^2 + \frac{dr^2}{1 - \frac{r_s}{r}} + r^2 d\Omega_2^2. \tag{1.1.2}$$

By fixing the angular direction and using the "zooming in" coordinates $r \to r_s(1 + \frac{\rho^2}{4r_s^2})$ and $t \to 2r_s\tau$, we can carry a small distance Taylor expansion about the horizon at $r = r_s$ with $\rho \ll r_s$, which applied to the Schwarzschild metric yields

$$ds^2 \approx -\rho^2\, d\tau^2 + d\,\rho^2. \tag{1.1.3}$$

This metric is manifestly Minkowskian under the change of coordinates

$$x^0 = \rho \sinh\tau, \quad x^1 = \rho\cosh\tau, \tag{1.1.4}$$

which, from elementary use of the chain rule applied to the above metric in "zoom in" coordinates, yields the flat space metric

$$ds^2 \approx -(dx^0)^2 + (dx^1)^2. \tag{1.1.5}$$

Here we see a manifest realisation of the principle of equivalence: an infalling object sees an environment that behaves locally like flat space, with no apparent force. It is effectively in free-fall, and there is nothing special locally from the objects own frame (before reaching the singularity). This is of course quite different from the point of view of an exterior observer, who could not retrieve signals past the horizon and would observe very peculiar effects like the apparent time dilation, redshifting and length contraction in a clearly warped geometry. This illustrates some of the paradoxical aspects of black holes.

It is also interesting to see that if we fix ρ, the coordinates in (1.1.4) have the form of a non-inertial frame with constant proper acceleration $1/\rho$, i.e. these are in fact Rindler coordinates. Physically, this amounts to say that if an object is placed near the horizon, it must have a constant acceleration to avoid falling in due to the gravitational attraction it experiences. The extreme case of this is when $\rho = 0$, meaning that the object needs infinite acceleration to leave the black hole, i.e. it has entered the horizon.

It is a curious property of non-inertial frames that the ground states of a QFT in an inertial frame appear thermal in the accelerating one. In particular, for accelerating observers describing the QFT, a density matrix with manifestly thermal behaviour arises. This result is known as the Unruh effect, and says that accelerating observers experience thermal baths which could in principle be measured by thermometers [3, 4].

Now, under the Euclidean continuation (i.e. Wick rotation) of the form $\tau = i\theta$, $x^0 = ix^0_E$, the new coordinate system describes the Cartesian plane in polar coordinates:

$$x^0_E = \rho\sin\theta \quad x^1 = \rho\cos\theta. \tag{1.1.6}$$

To see how this is related with thermal properties, consider the typical quantum partition function of the form

$$Z = \mathrm{Tr}[e^{-\beta H}], \tag{1.1.7}$$

where as usual, $\beta \equiv 1/T$ is the reduced temperature (in $k_B = 1$ units). By thinking of energy as the generator of time translations, it is clear that the argument of this formula has the form of a Euclidean time evolution. This formal similarity can be made precise when computing the partition function, where one can establish a direct mapping between the Euclidian time evolution along a circle of constant radius with total length $2\pi\rho$ and the temperature T in the thermal partition function, this is given by

$$\beta = \frac{1}{T} = \Delta \text{ Euclidean time} = 2\pi\rho. \tag{1.1.8}$$

By recalling that by fixing the radius, these coordinates also describe observers of constant proper acceleration $a = 1/\rho$ (see Fig. 1.1.1) and reinstating constants, we recover Unruh's result for the temperature of the thermal both seen by accelerating observers:

$$T_{\text{Unruh}} = \frac{\hbar}{k_B c}\frac{a}{2\pi}. \tag{1.1.9}$$

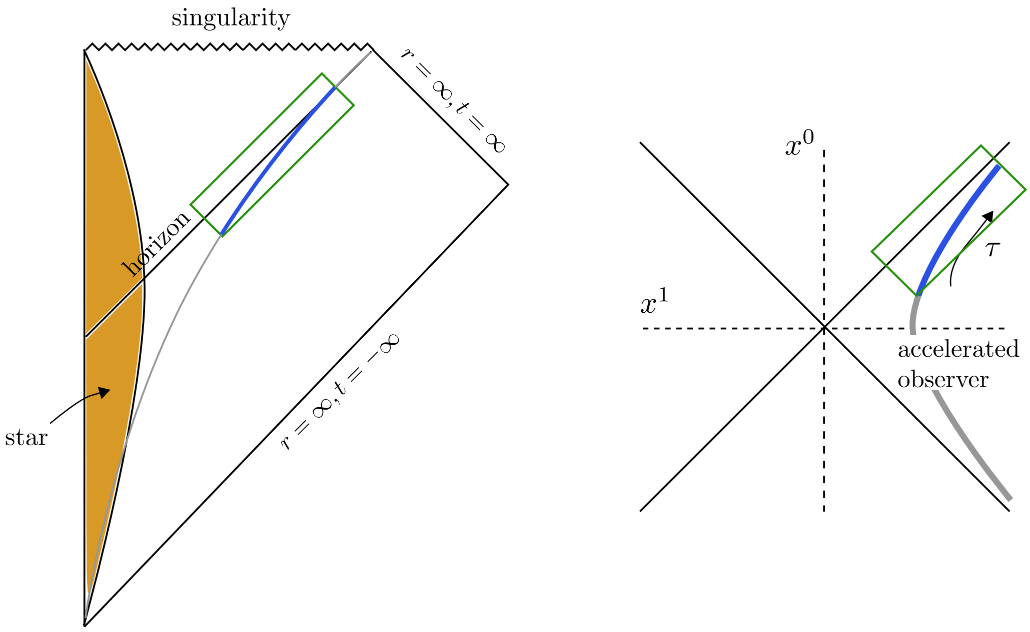


Image credit: [2]

Figure 1.1.1: In the left, the Penrose diagram of a black hole formed by gravitational collapse and in the right the worldlines of Rindler observers experiencing constant proper acceleration described by hyperboloids. We see how locally for static observers near the horizon, these two frames are equivalent.

This is the proper temperture experienced by observers at fixed positions near the horizon. To extend this further for distant observers for $r \gg r_s$, one can apply the Tolman relation, which accounts for the redshifting as one climbs away from the gravitational potential, this is given by [2, 5]

$$T(r)\sqrt{-g_{\tau\tau}(r)} = \text{constant}, \tag{1.1.10}$$

where by direct comparison with our Minkwskian metric and the form of the Unruh temperature, we see that the constant in dimensionless form is $1/(2\pi)$. It is then a simple exercise by changing back to Schwarzschild coordinates, that in the limit $r \gg r_s$ one obtains

$$T_H = \frac{1}{4\pi r_s}, \tag{1.1.11}$$

which by reinstating constants is indeed the Hawking temperature in (1.1.1).

1.2 Unitarity and the Page Curve

In another famous calculation by Hawking [6] the principle of unitarity, i.e. the fundamental notion that quantum systems should evolve according to deterministic equations of motion (like the Schrödinger equation) was called into question. This roughly corresponds to the question of whether information is destroyed when entering black holes.

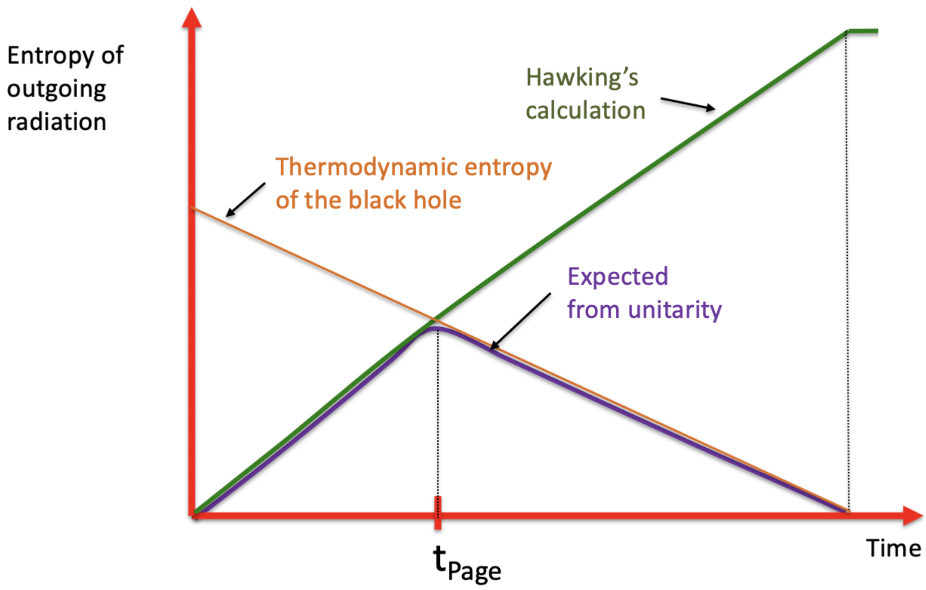


Image credit: [2]

Figure 1.2.1: The Page curve of Hawking radiation vs Hawking's prediction. According to the Page's reasoning, the entropy of the radiation should initially steadily increase due to the emission of effectively thermal states. It eventually saturates the Bekenstein bound at what is known as the Page time and remains obeying it thereafter until the black hole evaporates completely. Hawking's calculation conflicts with this, as while the early-time behaviour is similar, it eventually violates the Bekenstein bound and keeps increasing subsequently, reaching a high, constant valueafter evaporation. The fact that it is non-zero implies that the state is not pure, calling unitarity into question.

In particular Hawking showed that distant observers computing the thermodynamic entropy of radiation would exceed the classical bound that is allowed for the entropy, known as the Bekenstein bound and given by the area of the horizon A as

$$S_{\text{Bek}} = \frac{k_B c^3 A}{4 G_N \hbar}, \tag{1.2.1}$$

which we will derive in the next section.

The problem pointed by Hawking is that even if one starts in a pure state, the entropy of the radiation exceeds this bound and one seems to be forced to make the conclusion that the resultant state is mixed, meaning that some non-unitary process has erased information of the system, as it is no longer deterministic.

Under general quite general arguments, Page later argued how the entropy of such a radiation such look for unitarily evolving black holes and how this conflicted with Hawking's calculation [7]. This is known as the Page curve and it is shown in Fig. 1.2.1.

Much of the next sections will be devoted to show how this violation does not necessarily occur, as the latest calculations suggest that Page curves could arise naturally in gravitational systems like black holes.

1.3 Bekenstein-Hawking Entropy and the Holographic Principle

While Hawking's argument for the loss of unitarity is quite a convincing one, it has one crucial loophole. There is the implicit assumption that it is the thermodynamic entropy of the radiation that may be compared against the degrees of freedom described by the Bekenstein-Hawking entropy, so that once this bound is exceeded, these can no longer be purified with outgoing radiation states. This would mean that the system as a whole can no longer be pure, since it has a net entropy.

But to fully understand how the degrees of freedom of the interior and exterior black hole regions can be mapped with each other, it is necessary to develop an information-theoretic generalization of the entropy (which happens to equate with the thermal entropy only when maximized).

In 1973, when deriving his famous entropy formula for black holes, Bekenstein realized that there was a key distinction between the entropy described in his formula and the conventional thermal entropy implicit to Hawking's argument [8]:

> "It is then natural to introduce the concept of black-hole entropy as the measure of the *inaccessibility* of information (to an exterior observer) as to which particular internal configuration of the black hole is actually realized in a given case. At the outset it should be clear that the black hole entropy we are speaking of is *not* the thermal entropy inside the black hole."

The key relation between entropy and information is essentially that entropy is a measure of one's ignorance about a system's configuration in relation to some known bulk observable. This is clearly the case in the thermal case as exemplified by the Boltzmann formula, where entropy can be seen as the degeneracy of an ensemble in relation to thermodynamic observables like pressure, temperature, energy etc. The greater this number, the greater one's ignorance about the microscopic configurations of the system that lead to those observables, so we can extract from them less information about the system.

For example, as illustrated in [8], the entropy lost when compressing isothermically a gas in a closed container can be attributed to the reduced uncertainty in the position of the particles that constitute it, which significantly reduces the space of possible microstates.

There is no reason in principle why this type of entropy should be limited to thermodynamic systems, and in fact, the one used by Bekenstein to determine his formula came from information theory and is known as the Shannon entropy [9]:

$$S = -\sum_i p_i \ln p_i, \tag{1.3.1}$$

where we consider a set of possible outcomes, each with probability p_i. This entropy is maximized when the number of conditions on such probabilities is minimal, i.e. when they are all equal. If we consider such scenario by setting $p_i = 1/\Omega$ and interpret Ω as the number of microstates from a thermodynamic system, then we get

$$S = \ln \Omega, \tag{1.3.2}$$

which up to a choice of units is indeed the Boltzmann formula. It is in this sense that we can regard the thermal entropy as the maximization of the information entropy.

Bekenstein was motivated by classical theorems of black holes like Hawking's proof that the area of event horizons increases in time [10] and their close similarity with the second law of thermodynamics (as well as other related analogous behaviour with thermodynamic systems). This made him postulate that the entropy of a black hole should be proportional to its area, which accounts for the fact that event horizons shrink under evaporation, in violation of Hawking's original theorem. This part of what became known as the Generalized Second Law (GSL) [8, 11].

On dimensional grounds, he concluded that to make this entropy dimensionless it would be necessary to divide by the square of a fundamental length scale, the Planck length, which can be constructed as

$$l_P = \sqrt{\frac{\hbar G_N}{c^3}}. \tag{1.3.3}$$

The full expression with the remaining numerical factor was determined through an elegant thought experiment from a semiclassical information standpoint as follows (we will use $G_N = c = k_B = 1$ units). Define the minimal unit of information as the answer to a yes-or-no question, known as the bit, which trivially from Eq. (1.3.1) has an information content of $\ln 2$ (by which we mean the maximal entropy generated by the outcomes of a binary trial, such as a fair coin flip).

Then consider an observer outside the event horizon sending an infalling elementary particle with no internal structure and asking the question "does this particle exist?". Initially the answer is obviously yes, but once the particle enters the black hole, this becomes uncertain as the particle no longer becomes accessible beyond the event horizon.[2] This would imply that the black hole system has increased its information entropy by $\ln 2$ (one bit), which becomes the minimal increase in entropy (and thus horizon area) that may be gained from infalling objects.

[2] Technically, according to black hole complementarity, the particle would take an infinite time to reach the event horizon for an outside observer (even though it takes a finite proper time to cross it). But in practice, there will always be a finite time at which the time dilation and the red-shifting is so large that no signal can be sent from the particle to the observer, which suffices for our purposes.

The minimal increase in horizon area due to a spherical particle of mass μ and proper radius b (which is independent of the black hole parameters) is given by [8]

$$(\Delta A)_{\min} = 8\pi\mu b. \tag{1.3.4}$$

This is minimized by setting the particle radius to the (reduced) Compton wavelength $\hbar/\mu$,[3] provided that this is greater than the Schwarzschild radius ($r_s = 2\mu$), so that

$$b = \hbar/\mu > 2\mu, \tag{1.3.5}$$

meaning that the particle carries minimal information (no internal structure) while not converting itself into a microscopic black hole. This leads to the following relation for the entropy:

$$(\Delta S)_{\min} = \ln 2 = \frac{\mathrm{d}S(A)}{\mathrm{d}A}(\Delta A)_{\min} = 8\pi\hbar\frac{\mathrm{d}S(A)}{\mathrm{d}A}. \tag{1.3.6}$$

Integrating both sides gives

$$S(A) = \frac{\ln 2}{8\pi\hbar}A = \frac{\ln 2}{8\pi {l_P}^2}A. \tag{1.3.7}$$

A similar relation can be independently obtained from the Hawking temperature of a black hole (for an observer at infinity) [4]

$$T_H = \frac{\hbar}{8\pi M_{\mathrm{bh}}} \tag{1.3.8}$$

when combining it with the elementary thermodynamic relation

$$\frac{\mathrm{d}S}{\mathrm{d}E} = \frac{1}{T} \tag{1.3.9}$$

for a black hole with internal energy $E = M_{\mathrm{bh}}$ and horizon area $A = 4\pi {r_s}^2 = 16\pi M_{\mathrm{bh}}^2$, from which we get the famous black hole entropy formula

$$S(A) = \frac{A}{4\hbar} = \frac{A}{4{l_P}^2}. \tag{1.3.10}$$

Note the prefactor of 1/4 in Hawking's calculation, comparing this with Eq. (1.3.7), we see that Bekenstein's result has the exact same form and fundamental constants involved, but it is about an order of magnitude smaller. This perhaps illustrates the fact that Bekenstein's calculation was done from an information-theoretic perspective, while Hawking's comes straight from the thermodynamic relations as seen in Eq. (1.3.9). This is consistent with thermal entropy being necessarily greater than the information entropy,

[3] To give a quick motivation for this, from the uncertainty principle we have $\Delta p \geq \hbar/2\Delta x$, considering the relativistic energy $E^2 = p^2 + m^2$, we see that once the uncertainty in the energy due to the momentum exceeds the rest energy of the particle m, particle production could take place, leading to a measurement constraint. This bound is saturated when $\Delta x = \lambda/2 = \hbar/2m$, thus leading to the Compton wavelength.

[4] As an interesting coincidence, M_{bh} stands for the black hole mass, not Bekenstein-Hawking (thus the lack of capitalization). As pointed by other authors, it is also amusing that schwarz-schild means "dark shield" in German, who would have thought that a person with this name would play such a big role in understanding event horizons?

so that in this light, we may interpret Bekenstein's result as a lower bound of its thermal counterpart.

We see that the "Planck area" ${l_P}^2$ behaves like a fundamental unit of entropy, so that when computing the entropy one counts the number of such "microscopic tiles" that may be fitted in the area of the event horizon, leading to a finite number of degrees of freedom.

A remarkable fact about this formula is that it says that the entropy of a black hole does not scale with its volume (as one would expect from extensive quantities like entropy), but with its area, suggesting that its true degrees of freedom are encoded in its surface, not its volume.

This was the first realisation of what is now known as the holographic principle, which is the general idea that the information that describes a system is in some sense fundamentally encoded in its boundary (which has one less dimension). This is much like a hologram encoding a three-dimensional picture in an effectively two-dimensional surface, thus the name.

Bekenstein later conjectured an entropy bound for arbitrary regions of spacetime that precisely saturates the black hole entropy formula, implying that black holes are regions of maximal information, and thus that our universe is fundamentally holographic in that sense [12], this is known as the holographic entropy bound. The intuition behind this is that one cannot encode an arbitrary amount of information in a finite region of space, since this would require an indefinite amount of energy, which would eventually lead to black hole formation when the Schwarzschild limit is exceeded [13].[5] Based on Bekenstein's entropy bound, Susskind and t'Hooft argued that theories of quantum gravity should, in some well-defined sense, live in one dimension less than the usual $(d + 1)$-dimensional spacetime that they describe. This led to a more precise notion of the holographic principle and the original coinage of the term by Susskind [13, 14].

Although it hadn't been explicitly shown for some time that this kind of entropy was the result of internal microstates in black holes, this was proven in 1996 for a class of supersymmetric string theories that describe so-called BPS (Bogomol'nyi–Prasad–Sommerfield) black holes [15], strongly suggesting that this entropy has direct physical relevance. Also, as we will see in later sections, there are generalisations of this formula (as well as other instances of the holographic principle) that are being studied at the present moment (for which many pieces of independent, non-trivial evidence have been found) that have played a decisive role in understanding unitarity in black holes.

[5] Indeed, although finite, the number of degrees of freedom described by the entropy formula is enormous for astrophysical black holes. For example, a solar mass black hole has an entropy of order 10^{77}, while that of the Sun is $\sim 10^{60}$, which is much less. The entropy of Sagittarius A*, the black hole of circa 4 million solar masses at the centre of the Milky Way, is $\sim 10^{90}$. This is about equal to that of the CMB radiation, which has the dominant contribution to the entropy of the observable universe excluding black holes. Since they should be found in most galaxies, black holes are believed to be quite common, so it seems plausible that the vast majority of the information content in our universe comes from black hole microstates encoded in event horizons.

1.4 Fine-grained Entropy of Quantum Systems

Before we proceed further with our discussion on the entropy of black holes, it is necessary to further generalize our notion of entropy so that it can be applied to quantum systems.

Consider a pure (and for convenience orthonormal) state $|\Psi\rangle$ with density matrix $\rho = |\Psi\rangle\langle\Psi|$. Now consider the case where an observer is restricted (via some kind of partition or otherwise) to measure a subset of the Hilbert space of this system $\mathcal{H}_A$, the remainder (complement) space being $\mathcal{H}_B$, which is inaccessible to this observer. By tracing over $\mathcal{H}_B$, the observer in A sees an effective density matrix

$$\rho_A = \mathrm{Tr}_{\mathcal{H}_B}\rho. \tag{1.4.1}$$

This density matrix will in general describe a mixed state with entropy given by the von Neumann entropy formula as

$$S_A = -\mathrm{Tr}_{\mathcal{H}_A}[\rho_A \ln \rho_A]. \tag{1.4.2}$$

We see that the entropy of the system as a whole vanishes, since it is a pure state:

$$\begin{aligned} S_{\mathrm{tot}} &= -\mathrm{Tr}_{\mathcal{H}_{A\cup B}}[\rho\ln\rho] = -\mathrm{Tr}_{\mathcal{H}_{A\cup B}}\Big[\,|\Psi\rangle\langle\Psi|\sum_{n=1}^{\infty}c_n(|\Psi\rangle\langle\Psi|-1)^n\Big] \\ &= -\langle\Psi|\Psi\rangle\langle\Psi|\sum_{n=1}^{\infty}c_n(|\Psi\rangle\langle\Psi|-1)^n|\Psi\rangle = \sum_{n=1}^{\infty}(-1)^n c_n\langle\Psi|\left(|\Psi\rangle\langle\Psi|-1\right)|\Psi\rangle = 0, \end{aligned} \tag{1.4.3}$$

where we have done an elementary expansion of $\ln x$ about $x = 1$ imposing orthonormality.

The entropy seen from subsystem A is known as the entanglement or fine-grained entropy[6], the first terminology expresses the fact that the apparent degrees of freedom seen from subsystem A are purely due to the inaccessibility to subsystem B, so that each degree of freedom in one subsystem can be directly mapped to a "dual" one in the other one, making the state as a whole pure (i.e. "purifying" it), just like it would happen for a pair of entangled Bell states. This also makes intuitively obvious an important property found in these subsystems:

$$S_A = S_B, \tag{1.4.4}$$

which is necessary to establish this one-one mapping of degrees of freedom for a collectively pure bipartite system. This can be proved from the Schmidt decomposition [16], from which we can readily see that $\rho_A = \sum_i \lambda_i^2 |i_A\rangle\langle i_A|$ and $\rho_B = \sum_i \lambda_i^2 |i_B\rangle\langle i_B|$, where $|i_A\rangle, |i_B\rangle$, denote orthonormal states of subsystems A and B respectively, and λ_i are the Schmidt coefficients which act as identical eigenvalues for both ρ_A and ρ_B. Since the von Neumann entropy is purely determined by the eigenvalues of the density matrix (as it involves a trace), then it immediately follows that $S_A = S_B$.

[6] It should be noted that the term "entanglement entropy", despite its relevance and widespread use, is somewhat misleading in that this notion of entropy can be extended for mixed states too [4]. Note also that the definition of the entropy may be easily adapted for an arbitrary number of subsystems, but this is an unnecessary consideration for our purposes.

The term "fine-grained entropy" is used to emphasize the distinction (which we alluded to in the previous section) between this type of entropy and the thermal entropy, which in this context is known as the coarse-grained entropy. The reason why this distinction is so important is because the fine-grained entropy is compatible with unitary transformations of the form $\rho_A \to U\rho_A U^{-1}$, which can be seen by performing another elementary expansion of $\ln x$ and tracing over as:

$$\begin{aligned} S_A &\to -\mathrm{Tr}_{\mathcal{H}_A}[U\rho_A U^{-1}\ln(U\rho_A U^{-1})] = -\mathrm{Tr}_{\mathcal{H}_A}\Big[\sum_{n=1}^{\infty} c_n U\rho_A U^{-1}(U\rho_A U^{-1}-1)^n\Big] \\ &= -\mathrm{Tr}_{\mathcal{H}_A}\Big[\sum_{n=1}^{\infty} c_n U\rho_A U^{-1} U(\rho_A - 1)^n U^{-1}\Big] = -\mathrm{Tr}_{\mathcal{H}_A}[U\rho_A \ln \rho_A U^{-1}] = S_A, \end{aligned} \tag{1.4.5}$$

where we have used the invariance of trace under cyclic permutations.

This is a crucial requirement to overcome the information problem, as it means that this entropy (unlike the thermal one) is in general invariant under unitary evolution, so that if the system as a whole starts in a pure state, it will (under this class of transformations) remain so in agreement with unitarity.

We are now in the position to give a general notion of the thermodynamic entropy, which is as follows [2]: consider a system with density matrix ρ_0 that we are restricted to measure with a set of coarse-grained observables $\{A_i\}$ (i.e. simple observables which carry limited information of the system's configuration). The coarse-grained (thermal) entropy corresponds to the choice of ρ which maximizes the von Neumann entropy ($-\mathrm{Tr}[\rho \ln \rho]$) under the constraint that it yields the same expectations (i.e., observable outcomes) as the original density matrix so that $\mathrm{Tr}(\rho_{max} A_i) = \mathrm{Tr}(\rho_0 A_i)\ \forall\ i$. We also see here that the coarse-grained entropy is a measure of maximal ignorance about a system with a given a set of what can be regarded as bulk observables.

Another important property of the entropy is subadditivity, which says that

$$S_A + S_B \geq S_{A\cup B}. \tag{1.4.6}$$

This can be proven as follows [16]: Klein's inequality says that for two $n \times n$ positive-definite matrices ρ and σ and a differentiable convex function f, one can establish that

$$\mathrm{Tr}[f(\rho) - f(\sigma) - (\rho - \sigma) f'(\sigma)] \geq 0, \tag{1.4.7}$$

it can be easily seen that by setting $f(\rho) = \rho \ln \rho$[7] then $S(\rho) \leq -\mathrm{Tr}[\rho \ln \sigma]$, and by also setting $\rho = \rho_{A\cup B},\ \sigma = \rho_A \otimes \rho_B$, we have

$$\begin{aligned} S_{A\cup B} &\leq -\mathrm{Tr}_{\mathcal{H}_{A\cup B}}[\rho_{A\cup B}\ln(\rho_A \otimes \rho_B)] = -\mathrm{Tr}_{\mathcal{H}_{A\cup B}}[\rho_{A\cup B}(\ln \rho_A + \ln \rho_B)] \\ &= -\mathrm{Tr}_{\mathcal{H}_{A\cup B}}[\rho_{A\cup B}\ln \rho_A + \rho_{A\cup B}\ln \rho_B] = -\mathrm{Tr}_{\mathcal{H}_A}[\rho_A \ln \rho_A] - \mathrm{Tr}_{\mathcal{H}_B}[\rho_B \ln \rho_B] \\ &= S_A + S_B, \end{aligned} \tag{1.4.8}$$

where we have used the fact that when tracing over $\mathcal{H}_{A\cup B}$, only the elements that are traced either through A or B survive, [8] and applied the linearity of trace.

[7] Note that $f(x) = x \ln x$ is only convex between the interval $x \in [0, e]$, but the eigenvalues of the density matrix ρ are just probabilities (see Eq. (1.4.12)), which clearly must satisfy this bound.

[8] This can be seen explicitly by expanding the logarithm and applying orthogonality of states.

Comparing a system's behaviour with this property can be seen as a measure of the degree to which the entropy is extensive. In particular, for classical thermodynamic systems we have

$$S_{A\cup B} = S_A + S_B, \tag{1.4.9}$$

so we can see that the bound is saturated as a consequence of the entropy being maximal. On the other hand, for a pure state $S_{A\cup B} = 0$, which effectively saturates the bound from below, since the von Neumann entropy is a non-negative quantity.[9]

The notational similarity between expressions (1.3.1) and (1.4.2) is not a coincidence. Consider a density matrix ρ describing an effectively mixed state in a subsystem so that

$$\rho = \sum_i p_i \left|\psi_i\right\rangle \left\langle\psi_i\right|, \tag{1.4.10}$$

this yields the von Neumann entropy[10]

$$\begin{aligned} S = -\mathrm{Tr}[\rho \ln \rho] &= -\sum_i \left\langle\psi_i\right| \sum_j p_j \left|\psi_j\right\rangle \left\langle\psi_j\right| \ln\Big(\sum_k p_k \left|\psi_k\right\rangle \left\langle\psi_k\right|\Big) \left|\psi_i\right\rangle \\ &= -\sum_i p_i \left\langle\psi_i\right| \ln\Big(\sum_k p_k \left|\psi_k\right\rangle \left\langle\psi_k\right|\Big) \left|\psi_i\right\rangle = -\sum_i p_i \ln p_i, \end{aligned} \tag{1.4.11}$$

where the last step can again be understood by making an elementary expansion and imposing orthonormality. Thus, we see how the Shannon entropy follows naturally from this prescription. In fact, this can be seen as the obvious quantum-mechanical generalization of the Shannon entropy, since from the obvious property

$$\rho \left|\psi_i\right\rangle = p_i \left|\psi_i\right\rangle, \tag{1.4.12}$$

we see that the density matrix is none other than the operator whose egienvalues yield the Shannon probabilities, meaning that the difference between the two prescriptions really is almost purely notational. The quantum prescription arises due to the necessity of a formalism to compute such probabilities, but is still entirely consistent with the original formula, which is once more a testament to its wide applicability.

It is worth noting that there is a deep similarity between the bipartite quantum systems we have discussed and black holes. In particular, an event horizon acts like a partition, so we may regard system A as an outside observer and system B as the black hole interior, which further illustrates the relevance of the fine-grained entropy as the kind of entropy originally described by Bekenstein.

This analogous behaviour can be taken to the next level by obtaining an area law for the entropy which need not to apply to black holes only, as realised by Srednicki [17]. The argument was, again, quite simple. Consider a free scalar quantum field theory with a ground state density matrix $\rho = \left|0\right\rangle \left\langle 0\right|$. Now consider dividing the space into a spherical partition of radius R. Define A to be the exterior region and B the interior. The density

[9] For a proof of this and additional properties of the von Neumann entropy, see [16].

[10] Where once more we have assumed orthonormality of states, which is more or less the quantum equivalent of the Shannon probabilities $\{p_i\}$ describing mutually exclusive events.

matrix of the exterior region A depends only on the degrees of freedom of the region itself, as the complementary region B's degrees of freedom are eliminated with the partial trace as in Eq. (1.4.1) (for the converse argument, simply make the replacement $A \leftrightarrow B$).

A naive argument would be to suppose that the entropy in each region is determined by its volume, but from Eq. (1.4.4), we know this can't be the case as both entropies should be equal (but the volumes obviously are not). The only geometric commonality between these two regions is in fact the area of the partition, which is $4\pi R^2$.

To make the entropy dimensionless, we resort to the two other dimensionful parameters that are necessary to define this QFT, the ultraviolet and infrared cut-offs M and μ respectively. But we know that the long wavelength physics should not affect the inner region B, so it gets discarded. Therefore, on dimensional grounds, one obtains that

$$S = \kappa M^2 A = \frac{\kappa}{l^2} A, \tag{1.4.13}$$

where κ is a numerical factor to be determined and we have made a suggestive parametrisation involving a length scale $l \equiv 1/M$ to allow for direct comparison with the black hole entropy formula (1.3.10). Srednicki then proceeded to show specific instances of this relation in typical quantum systems, further confirming the idea and its very close relationship with black hole entropy.

2 The AdS/CFT Correspondence

It is now time to introduce what is arguably the most successful (and certainly the most cited) realisation of the holographic principle, the anti-de Sitter / conformal field theory (AdS/CFT) correspondence. This is also known as the gauge/gravity duality, which captures the fact that it is a correspondence between theories of gravity and gauge theories which applies beyond the simpler, more studied cases of AdS spaces and CFTs. It was originally conjectured by Maldacena in 1998 [18] from a string theory standpoint,[11] but has since been generalised further and found support across a number of different fields beyond string theory, converting itself into a vast multidisciplinary subject.

In our context, as we will see, it has played a foundational role in generalizing the area formula for the entropy, which is necessary in order to reproduce Page curves, yielding a potential theoretical resolution to the information problem in black holes.

In essence, the AdS/CFT correspondence establishes a close connection between theories of gravity in D-dimensional asymptotically AdS spacetimes and a class of strongly coupled QFTs (the CFTs) living far in the D-1 dimensional boundaries of such spacetimes, known as conformal boundaries. The idea is that the CFTs encode all the information of those theories of gravity in AdS spacetime in a precise and powerful way: one may establish a dictionary between calculations/experiments done in one theory vs the other, which includes key quantities like correlation functions, and of course, entropy.

The reason why this is useful is because quantum gravity is a notoriously difficult subject, while QFT is relatively well understood, so with this correspondence we can probe non-trivial behaviour in theories of quantum gravity through the lens of comparatively simple QFTs. In the context of black holes, it is strongly suggestive of the fact that any black hole living in such AdS spacetimes would necessarily comply with unitarity, as its behaviour could be mapped directly to a QFT that is manifestly unitary. This has been shown to be true for specific instances of the duality, and has also led many people, including Hawking, to move to the pro-unitarity camp when it comes to resolutions of the information problem.

Of course, these spacetimes are quite exotic and don't correspond to our universe, where astrophysical black holes live. But the beauty of these arguments is that by studying these toy models, one can derive general conclusions about theories of gravity which are not dependent on such environments, which has been the case with the gravitational entropy formulas that will be presented in later sections.

2.1 Anti-de Sitter Space

Let us start by examining the left hand side of the correspondence. AdS spacetime is the negative cosmological constant, maximally-symmetric[12] solution to the Einstein field equations in the vacuum (i.e. with a vanishing stress-energy tensor). The other two being

[11] In particular, in the context of establishing a dual relationship between $\mathcal{N} = 4$ supersymmetric Yang-Mills theory in $3 + 1$ dimensions and type IIB string theory living on $AdS_5 \times S^5$.

[12] These are homogeneous and isotropic spacetimes that encode the maximum number of symmetries allowed for their dimension, these isometries are described by their Killing vector fields.

the flat Minkowski (vanishing cosmological constant) and de Sitter (positive cosmological constant) spacetimes.[13]

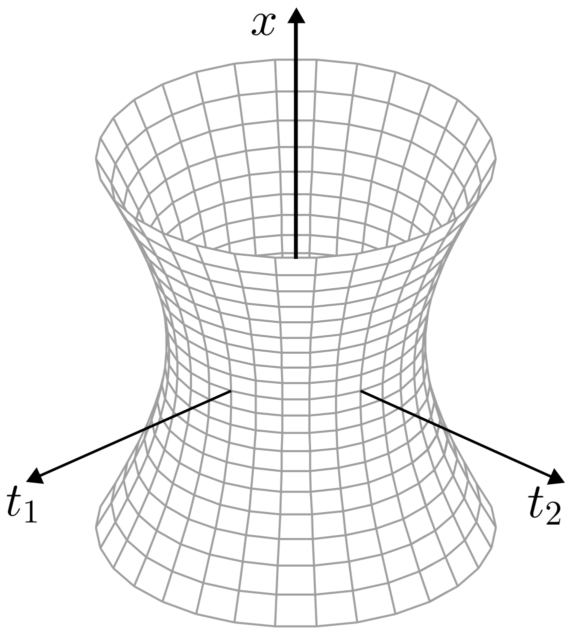


Image credit: Based on "Quadriken-7.svg" by Ag2gaeh, licensed under CC BY-SA 4.0.

Figure 2.1.1: A depiction of the hyperboloid surface that defines $(1 + 1)$-dimensional AdS spacetime in a $(1+2)$-dimensional flat ambient spacetime. There is one vertical space dimension and two timelike ones in the plane of rotation symmetry, these are represented with axes. The flat space embedding allows us to define the geometry of this space, while also allowing for a clear illustration of its hyperbolic curvature.

Geometrically, AdS has a constant negative Ricci scalar curvature which is associated with a hyperbolic geometry (de Sitter on the other hand has constant positive curvature, which is associated with a spherical geometry). While these associations come originally from Euclidean geometry, they can be extended by embedding a $(p+q)$-dimensional AdS spacetime in $\mathbb{R}^{p,q+1}$ (flat spacetime with p space and $q + 1$ time dimensions), with the metric of the embedding space being

$$ds^2 = \sum_{i=1}^{p} dx_i^2 - \sum_{i=1}^{q+1} dt_i^2. \tag{2.1.1}$$

The embedded surface that defines the AdS_{p+q} space then takes the form of a generalized hyperboloid with radius of curvature L as

$$\sum_{i=1}^{p} x_i^2 - \sum_{i=1}^{q+1} t_i^2 = -L^2, \tag{2.1.2}$$

[13] By and large, our universe appears to have a small positive cosmological constant, corresponding to the geometry of a nearly-flat de Sitter spacetime.

which has a manifest $SO(q+1,p)$[14] symmetry that describes the full set of isometries in AdS, as this expression defines a constant generalized inner product of vectors. An illustration of this is shown in Fig. 2.1.1 for $(1+1)$-dimensional AdS spacetime.

Note that despite the counter-intuitive notion of a universe with two time dimensions, these are more of an artefact of the embedding, as creatures living in this AdS universe would only see one time dimension that wraps around the hyperboloid. Note also that the time dimension is compact (there are closed time-like curves), this can be "unwrapped" by taking the universal cover, so that for our purposes we will define the time dimension to be unbounded in this way for AdS space.

For the remainder of our discussion on AdS, we shall focus on AdS$_3$, i.e. AdS in $2+1$ dimensions,[15] then Eq. (2.1.1) and Eq. (2.1.2) reduce to

$$ds^2 = dX^2 + dY^2 - dU^2 - dV^2 \tag{2.1.3}$$

and

$$X^2 + Y^2 - U^2 - V^2 = -L^2 \tag{2.1.4}$$

respectively, where X, Y and U, V are spacelike and timelike dimensions (respectively).

It can be easily checked that solutions to Eq. (2.1.4) may be parametrized in the global coordinates $\rho \geq 0$ and $\tau, \theta \in [0, 2\pi]$[16] as

$$\begin{aligned} U &= L\cosh\rho\cos\tau, \\ V &= L\cosh\rho\sin\tau, \\ X &= L\sinh\rho\cos\theta, \\ Y &= L\sinh\rho\sin\theta. \end{aligned} \tag{2.1.5}$$

Substituting this result into Eq. (2.1.3) gives the AdS$_3$ metric

$$ds^2 = L^2\Big[-\cosh^2\rho\, d\tau^2 + d\rho^2 + \sinh^2\rho\, d\theta^2\Big], \tag{2.1.6}$$

where we have straightforwardly applied the chain rule so that

$$\begin{aligned} dU &= L(\sinh\rho\cos\tau\, d\rho - \cosh\rho\sin\tau\, d\tau), \\ dV &= L(\sinh\rho\sin\tau\, d\rho + \cosh\rho\cos\tau\, d\tau), \\ dX &= L(\cosh\rho\cos\theta\, d\rho - \sinh\rho\sin\theta\, d\theta), \\ dY &= L(\cosh\rho\sin\theta\, d\rho + \sinh\rho\cos\theta\, d\theta). \end{aligned} \tag{2.1.7}$$

Making the so-called static coordinates reparametrization $r = \sinh\rho$, we may re-express the metric into a more familiar form as

$$ds^2 = L^2\Big[-(1+r^2)\, d\tau^2 + \frac{dr^2}{1+r^2} + r^2\, d\theta^2\Big], \tag{2.1.8}$$

[14] Here we have used the convention where the timelike dimension is written first. Some authors do the converse, but that is purely notational in this context, as both cases refer to the same Lorentz symmetry group with metric signature $(q+1,p)$.

[15] This is the simplest case that captures the key properties of AdS which may be easily generalized to further dimensions, it also happens to be the one that was originally studied for generalizing the entropy formula for black holes discussed in the next section.

[16] Where again, it should be understood that the universal cover may be taken to de-compactify the timelike dimension τ, as it is usually done for cases of physical interest.

which happens to be especially relevant for studying black holes in AdS. We see that this rendition of the metric approaches Minkowski space in the limit $r \to 0$ and that τ is the proper time for an observer at $r = 0$, meaning that AdS is a Lorentzian manifold. The metric is also clearly spherically symmetric, i.e. it has an $SO(2)$ symmetry as expected from its full symmetry group $SO(2,2)$ (or rather, its universal cover for non-compact τ).

Insight may be gained by compactifying the space as $r = \tan\chi$, $\chi \in [0, \pi/2)$, obtaining the metric

$$ds^2 = \frac{L^2}{\cos^2\chi}\Big[-d\tau^2 + d\chi^2 + \sin^2\chi\, d\theta^2\Big]. \tag{2.1.9}$$

This allows us to construct a Penrose diagram of AdS spacetime, as seen in Fig. 2.1.2.

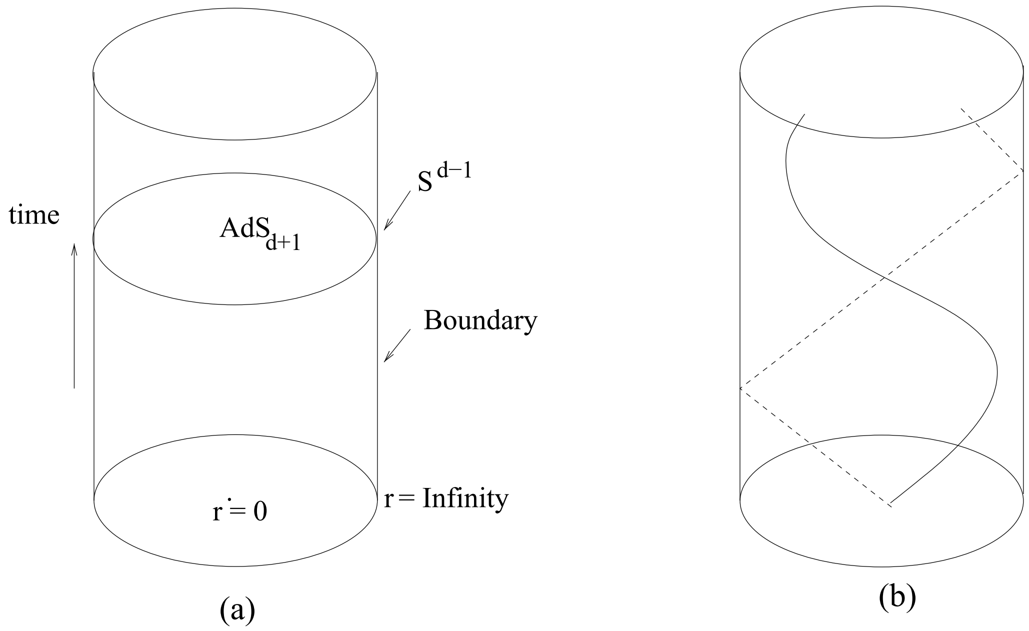


Image credit: [19]

Figure 2.1.2: Penrose diagram of AdS$_{d+1}$ spacetime. Time is measured as the proper time for a free-falling observer at $r = 0$. Only the radial coordinate has been comformally compactified, leading to a timelike-infinite cylindrical boundary (where CFTs are found) for AdS$_3$, or more generally an $\mathbb{R} \times S^{d-1}$ surface. Each horizontal slice represents a Cauchy slice of such universe, these can be stacked in time to form a cylinder as illustrated in (a). In (b) we see timelike wordlines of massive particles which are restricted from reaching the boundary in a finite time, while null geodesics can with reflecting boundary conditions.

We see that timelike trajectories of massive particles take infinite proper time to reach the boundary at $\chi = \pi/2$, while a null geodesic, i.e. $ds^2 = 0$, takes a finite time[17]

$$\Delta\tau = 2\int_0^{\pi/2} d\chi = \pi, \tag{2.1.10}$$

[17] Assuming reflecting boundary conditions, meaning that null geodesics “bounce back” at the time-like boundary of the space.

thus confirming the timelike nature of the boundary. In particular, AdS behaves like a box in that particles within its boundaries are confined and only massless particles are ever able to reach it and bounce back. This is not unlike the notion of a particle in a box potential, as there is an increasing gravitational potential away from the origin at $r = 0$ that effectively confines particles in this way. Specifically, slow moving particles experience a gravitational potential $V \sim \sqrt{-g_{\tau\tau}}$ [19], so that for $r \ll 1$, Taylor expanding gives $V \sim L\sqrt{1+r^2} \approx L + \frac{L}{2}r$, which is the same form as the harmonic potential. Thus, we see that particles that are perturbed away from the origin will execute harmonic-like motion about it due to this.

Note also that, since AdS is a Lorentzian manifold, there is nothing special about the $r = 0$ point, as it is just comes from some arbitrary free-falling reference frame. As in usual Minkowski space, one can always create boosts where the AdS universe is seen in the rest frame of any particle whose worldline makes a timelike geodesic, so there is no universal center in AdS (even though it may seem like so from the Penrose diagram in Fig. 2.1.2).

Let's now consider another useful set of coordinates, the Poincaré coordinates (t, x, z), which are parametrized as

$$t = \frac{\sin\tau}{\cos\tau + \sin\chi\cos\theta},$$

$$x = \frac{\sin\theta\sin\chi}{\cos\tau + \sin\chi\cos\theta}, \tag{2.1.11}$$

$$z = \frac{\cos\chi}{\cos\tau + \sin\chi\cos\theta},$$

where χ is the same conformal coordinate we used previously.

Again, by straightforwardly applying the chain rule as in (2.1.7), one gets the metric

$$ds^2 = \frac{L^2}{z^2}[-dt^2 + dx^2 + dz^2]. \tag{2.1.12}$$

We see that this is the same as the Minkowski metric up to a non-constant overall factor. We also see that at the boundary where $\chi = \pi/2$, then $\cos\chi = 0$, so $z = 0$.[18] An interesting peculiarity of this rendition of the metric is that as $z \to \infty$ the metric becomes null (it vanishes), meaning that there is a horizon in this limit.

Note that for constant z this metric truly reduces to that of $\mathbb{R}^{1,1}$ Minkowski space (up to a simple choice of units), which is precisely the canonical environment where QFTs are typically defined (it is also at constant $z = 0$). Note however that AdS is not globally flat in its boundaries, just locally, as it is a Lorentzian manifold.

Poincaré coordinates effectively zoom into the boundaries and reveal this flat structure, but they do not capture the full spacetime, which is curved. So that while the

[18]Note that the metric diverges at the boundary, but this is to be expected, as the "boundary" only exists because of the conformal compactification, so it is really infinite in extent. This behaviour is also observed in the other renditions discussed previously, but it is somewhat less evident because there is no denominator that vanishes.

conformal boundary is locally flat, this does not contradict the fact that CFTs live in the cylinder $\mathbb{R} \times S^1$ defining such boundary, which is not globally flat.[19]

This metric also has a manifest scaling symmetry under the transformation

$$(t, x, z) \to \lambda(t, x, z), \tag{2.1.13}$$

which is part of the conformal symmetry group, meaning that such class of transformations in the boundary create mappings to points that are also in the boundary, which is a necessary property for the CFTs to obey. In fact, the conformal group is isomorphic to the $SO(2,2)$ symmetry acting on the boundary. In this case, the rescaling symmetry (2.1.13) of the metric (2.1.12) acts as a dilatation on the conformal boundary. This is one aspect of the correspondence: there is a direct mapping between the symmetry generators in the AdS bulk and those of the conformal boundary.

2.2 Conformal Field Theory

CFTs are basically QFTs that are invariant under transformations of the conformal group, which much like the so-called conformal transformations for Penrose diagrams,[20] are a class of transformations that preserve angles and inner products of vectors, but not necessarily lengths. For CFTs, these symmetries comprise the Poincaré group (boosts + rotations + translations), dilatations, and the so-called special conformal transformations, which take the form

$$x^\mu \to \frac{x^\mu + a^\mu x^2}{1 + 2x \cdot a + a^2 x^2}, \tag{2.2.1}$$

although this latter symmetry is less relevant for our purposes.

CFTs have a number of characteristic properties. The simplest case in 3+1 dimensions is the free massless scalar field with action

$$I[\phi] = -\frac{1}{2} \int d^4x \sqrt{-g}\, g^{\mu\nu} \partial_\mu \phi\, \partial_\nu \phi. \tag{2.2.2}$$

We see that this action is invariant under dilatations of the form

$$\begin{aligned} x^\mu &\to \lambda x^\mu, \\ \phi(x) &\to \lambda^{-1} \phi(x). \end{aligned} \tag{2.2.3}$$

A characteristic common to all CFTs is the existence of what are known as primary operators. The special property about these is that they transform in a relatively simple way under conformal transformations. These operators comprise all local operators that commute with the special conformal generators at $x = 0$. Primary operators transform under dilatations as

$$\mathcal{O}(x) \to \lambda^{-\Delta} \mathcal{O}(x), \tag{2.2.4}$$

[19] Interestingly, there has been recent progress in establishing a correspondence similar to AdS/CFT but with asymptotically flat spacetimes (AFS) [20, 21, 22], meaning that the boundary CFTs of such theories truly live in Minkowski space and not in the more exotic conformal boundaries of asymptotic AdS spacetimes. This new field is known as celestial holography.

[20] These two terminologies are analogous but not necessarily the same.

where the exponent Δ is known as the conformal dimension. The conformal dimension is real and positive for unitary CFTs and equal or greater than $(d-2)/2$ for primary operators that are also scalars. Derivatives of primary operators are known as descendants, taken to the nth degree they have a conformal dimension $\Delta \to \Delta + n$, but will in general not be primary operators.

One useful property of primary operators is that their time ordered two-point correlation functions (which encode potentially useful physical quantities like scattering amplitudes) take a simple form. It is not difficult to see that in order to respect the property (2.2.4) and be Poincaré invariant, such correlation functions should (up to an overall constant) take the form[21]

$$\langle\Omega| T\mathcal{O}(x,t)\mathcal{O}(0,0)|\Omega\rangle = \left(\frac{1}{-t^2+|x|^2+i\epsilon}\right)^{\Delta}. \tag{2.2.5}$$

CFTs have another important property. According to AdS/CFT, the CFTs that are dual to quantum gravity in the AdS$_{d+1}$ bulk live in an $\mathbb{R}\times S^{d-1}$ "cylinder" with metric[22]

$$ds^2 = -d\tau^2 + d\Omega_{d-1}^2, \tag{2.2.6}$$

which under the Euclidean time continuation $\tau_E \equiv i\tau$ (i.e. a Wick rotation) becomes the usual Euclidean cylinder with metric

$$ds^2 = d\tau_E^2 + d\Omega_{d-1}^2. \tag{2.2.7}$$

This cylinder can be mapped to the Euclidean plane under the substitution

$$r = e^{\tau_E} \quad \Longleftrightarrow \quad \tau_E = \ln r, \tag{2.2.8}$$

so that the metric becomes becomes conformally equivalent to the Euclidean plane $\mathbb{R}^d$ as

$$ds^2 = d(\ln r)^2 + d\Omega_{d-1}^2 = \frac{1}{r^2}[dr^2 + r^2\, d\Omega_{d-1}^2] \simeq dr^2 + r^2\, d\Omega_{d-1}^2, \tag{2.2.9}$$

where we have exploited the fact that CFTs are invariant under the Weyl transformation multiplying the metric by the overall $1/r^2$ factor. Note in particular that dilatations in the Euclidean plane are equivalent to time translations in the cylinder, since from (2.2.8)

$$r \to cr \quad \Longleftrightarrow \quad \tau_E \to \tau_E + \ln c. \tag{2.2.10}$$

This relationship allows for a mapping between the eigenstates of the Hamiltonian in the cylinder (the generator of time translations) and local operators at the origin of the plane, which comprises primary operators that commute with the generators of the dilatations.

To see this, in analogy with wavefunctions in quantum mechanics, let's characterize states by wavefunctionals in Cauchy slices (i.e. constant time). For a CFT in the Euclidean cylinder, an initial wavefunctional of a field at position σ at time τ_i, $\Psi_i[\phi_i(\sigma),\tau_i]$, evolves to a final wavefunctional according to the path integral

$$\Psi_f[\phi_f(\sigma),\tau_f] = \int \mathcal{D}\phi_i \int_{\phi(\tau_i)=\phi_i}^{\phi(\tau_f)=\phi_f} \mathcal{D}\phi\; e^{-I[\phi]}\; \Psi_i[\phi_i(\sigma),\tau_i]. \tag{2.2.11}$$

[21]Where we have used the usual $i\epsilon$ prescription to deal with branch cuts.

[22]Where unlike in previous definitions of the time dimension τ, length units have been implicitly absorbed, as there is no factor of L^2.

The equivalent computation for a CFT in the Euclidean plane when integrating over the annulus between the radial coordinates $r_f > r_i$ with fixed boundary conditions ϕ_i, ϕ_f is given by

$$\Psi_f[\phi_f(\sigma), r_f] = \int \mathcal{D}\phi_i \int_{\phi(r_i)=\phi_i}^{\phi(r_f)=\phi_f} \mathcal{D}\phi \; e^{-I[\phi]} \; \Psi_i[\phi_i(\sigma), r_i]. \tag{2.2.12}$$

Note that in the first case (2.2.11) the wavefunctional over the initial state boundary condition at τ_i acts like a weighting factor on the path integral, while on the second case (2.2.12) the weighting factor is over the inner circle at r_i.

Now let's consider what happens in the latter case when $r_i \to 0$ so that we integrate over the whole disk up to $r_f = r$. The inner circle shrinks to a point at the origin and we are left with something that acts on the path integral which depends only on such point. This exactly matches the definition of a local operator at the origin $\mathcal{O}(0)$, and due to the conformal invariance, this establishes a bijective map between states in the CFT found in the cylinder and local operators in the plane as

$$\Psi_f[\phi_f, r] = \int^{\phi(r)=\phi_f} \mathcal{D}\phi \; e^{-I[\phi]} \; \mathcal{O}(0), \tag{2.2.13}$$

this important relationship is known as the state/operator correspondence and it is a general property completely independent from AdS/CFT. Fig. 2.2.1 summarises this with an illustration. Note that from (2.2.8), $\tau_E \to -\infty$ as $r \to 0$ and $\tau_E = 0$ when $r = 1$.

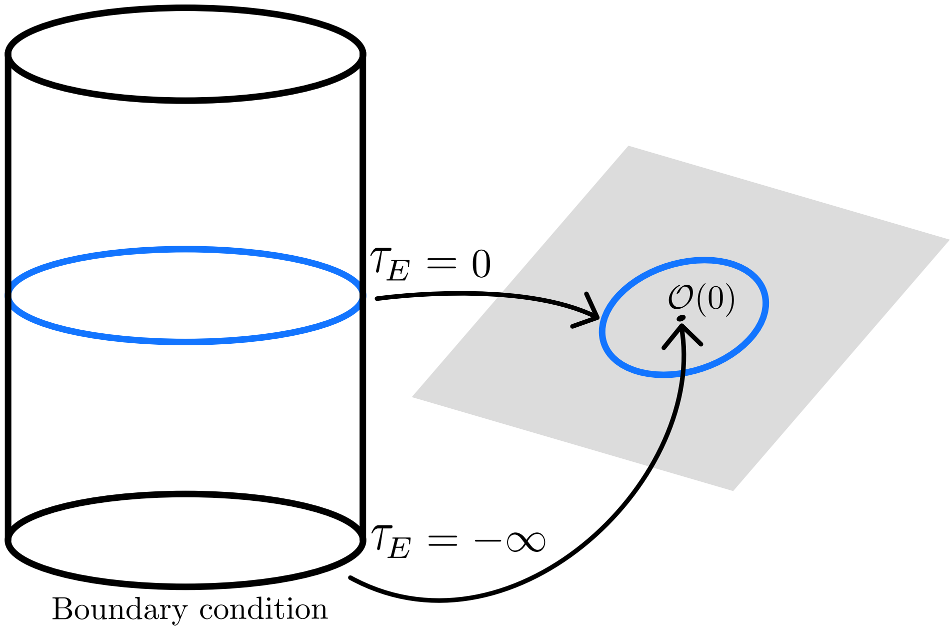


Image credit: Based on [23].

Figure 2.2.1: The state/operator correspondence. A local operator at the origin of the Euclidean plane $\mathbb{R}^d$ gets bijectively mapped to an eigenstate of energy Δ in the infinite past of a CFT in the Euclidean cylinder of $\mathbb{R} \times S^{d-1}$, where Δ is the operator's conformal dimension. The unit circle state in $\mathbb{R}^d$ corresponds to one of zero time in the cylinder.

2.3 Statement of the AdS/CFT Correspondence

We are now in the position to articulate what the AdS/CFT correspondence says more precisely [4, 24]:

> Any conformal field theory in $\mathbb{R} \times S^{d-1}$ with metric (2.2.6) can be mapped directly to and interpreted as a theory of quantum gravity in asymptotically $AdS_{d+1} \times M$, for some compact manifold M.

Although not necessarily the case, M is trivial for our purposes.

Asymptotically AdS basically means that the spacetime tends to AdS in the limit of approaching the boundary (which should be the only one and timelike) as $r \to \infty$.[23] This is much like say, how the Schwarzschild geometry approaches Minkowski space in the same limit. The reason why this is important is because plain AdS everywhere would be a relatively trivial and featureless environment for a theory of quantum gravity, so that more interesting realisations of AdS/CFT are obtained with perturbations of AdS space which nevertheless satisfy the asymptotic requirement.

The above can be regarded as a sensible definition of theories of quantum gravity living in AdS spaces in terms of better understood CFTs. However, it should be emphasized that AdS/CFT is not merely true by definition, as there are several non-trivial pieces of evidence that point to this conjecture being true, notwithstanding that it hasn't been rigorously proven generally, which is why it remains a conjecture [19].[24]

This correspondence can be seen as a dictionary that maps all the key relations between the physics of the bulk theory in AdS and the CFT at the boundary in essentially every way, so that both theories describe the same thing in a different language. So although the above statement of the correspondence may seem somewhat ambiguous in that it doesn't say specifically what it is that is being mapped or interpreted, it really isn't: any physical aspect that characterises one theory on one side must have an equivalent dual in the other one. For example, we already discussed at the end of Section 2.1 that both theories share the same $SO(2,2)$ isometry in $\mathrm{AdS}_3/\mathrm{CFT}_2$, which can be generally extended to $SO(2,d)$ for $\mathrm{AdS}_{d+1}/\mathrm{CFT}_d$, this is one of the many entries of the dictionary.

Another such entry which is related to such isometries is that both sides should have the same Hamiltonian, this also involves thermodynamic quantities that depend on it such as partition functions, free energies etc., which can all be defined from the CFT. Gauge theories in the boundary with global symmetries will also have dual conserved charges in the AdS bulk side due to the mapping of symmetries and Noether's theorem.

There is an important entry known as the extrapolate dictionary, which describes how bulk fields approach the behaviour of conformal primary operators with conformal dimension Δ from the CFT in the limit of reaching the boundary as [4, 26]

$$\lim_{r \to \infty} r^{\Delta} \phi(t, r, \Omega) \equiv \mathcal{O}(t, \Omega). \tag{2.3.1}$$

[23] For a rigorous definition, see [25].

[24] This means that AdS/CFT is thought to give genuine insights on what it would mean to have a consistent theory of quantum mechanics and gravity, it is not merely a string of logical propositions (although it does have a lot to do with strings).

For the more general case of expectations involving operator products, each bulk field's time must match that of its dual CFT operator.[25] This allows for direct comparison between key physical quantities such as correlation functions leading to scattering amplitudes. To see an explicit example of this consider the free massive scalar field in the AdS bulk, which has a time-ordered two-point correlation function [4, 27]

$$\langle\Omega|\, T\phi(x)\phi(x')\,|\Omega\rangle = (2\Delta-d)^{-1}2^{-2\Delta}\pi^{-d/2}\frac{\Gamma(\Delta)}{\Gamma(\Delta-d/2)}u^{\Delta}F\Big(\Delta,\Delta+\frac{1-d}{2},2\Delta-d+1,u\Big), \tag{2.3.2}$$

where

$$u=\frac{2}{1+\cosh\ell}, \tag{2.3.3}$$

ℓ being the length of the geodesic connecting the points x and x' in AdS and F a hypergeometric function. The conformal dimension is given by the mass of the scalar field as

$$\Delta=\frac{d}{2}+\frac{1}{2}\sqrt{d^2+4m^2}. \tag{2.3.4}$$

By expressing the geodesic in the compactified global coordinates used in the metric (2.1.9), one obtains[26]

$$u=\frac{1}{1-rr'\cos\alpha+\sqrt{(1+r^2)(1+r'^2)}\cos[(t-t')(1-i\epsilon)]}, \tag{2.3.5}$$

where α is the angular separation between the two points with respect to the origin. From Eq. (2.3.1) one then obtains that for the corresponding CFT correlation function

$$\langle\Omega|\, T\mathcal{O}(\alpha,t)\mathcal{O}(0,t')\,|\Omega\rangle \propto \lim_{r,r'\to\infty}(rr'u)^{\Delta}=\left(\frac{1}{\cos[(t-t')(1-i\epsilon)]-\cos\alpha}\right)^{\Delta}. \tag{2.3.6}$$

By setting $t'=0$ and taking $t,\alpha\ll 1$, so that we approach the Minkowskian limit of the manifold with two closely separated points at the boundary, then[27]

$$\langle\Omega|\, T\mathcal{O}(\alpha,t)\mathcal{O}(0,0)\,|\Omega\rangle \propto \left(\frac{1}{-t^2+\alpha^2+i\epsilon}\right)^{\Delta}, \tag{2.3.7}$$

which is consistent with the expected form from (2.2.5).

Finally, and most importantly for our purposes, there is another entry in the dictionary: the entropy. The conformal boundary encodes the same number of degrees of freedom and thus entropy as the AdS bulk, behaving much like the boundary of an entangled system. This will be discussed in the following section.

[25] Other choices would correspond to CFTs living in geometries other than the AdS cylinder in (2.2.6).

[26] Geodesics in AdS are discussed in Section 3.2, where they turn out to be especially relevant for entropy calculations in AdS_3/CFT_2. Note that the $i\epsilon$ prescription is necessary to pick the right branch.

[27] Where here we have simply expanded $\cos x\approx 1-\frac{x^2}{2}$ and absorbed a factor of t into the $i\epsilon$ term.

3 A General Expression for the Gravitational Entropy

While Bekenstein-Hawking (BH) entropy has played a foundational role in understanding degrees of freedom and unitarity in black holes, in order to fully understand this, one needs to fully reconcile it with the von Neumann entropy prescription. In particular, there are significant corrections to the BH entropy that arise from a full quantum treatment leading to the fine-grained version.

As it turns out, this is a crucial consideration to account for in a unitary treatment of black holes that would resolve the information problem. The BH entropy expression taken at face value still can't account for the apparent degrees of freedom arising from the Hawking radiation to exterior observers.

Motivated by AdS/CFT, in this section we will review how this general expression for the entropy is obtained.

3.1 Counting Degrees of Freedom in AdS/CFT

A natural sanity check of the AdS/CFT correspondence is two compare the degrees of freedom of the conformal boundary with those of the AdS bulk. Of course, when one talks about degrees of freedom, one implicitly talks about entropy, and as we hopefully made clear in Section 1.2, information is also implicit to this. So it is certainly necessary for the correspondence to be realised to include entropy as one of the entries in the dictionary.

But how can a bulk have the same amount of entropy as its surrounding boundary with one less dimension? This is another version of the holographic principle and its counter-intuitive nature in relation to our notions on the extensiveness of entropy. In fact, a thermodynamic treatment does lead to some contradictions, which would only seem to support more scepticism.

Consider a gas of bosons in the conformal boundary with an effective temperature T in the thermodynamic limit when $T \gg 1$ (compared to the radius of curvature), the entropy of such a gas scales as $S \sim V_{d-1} T^{d-1}$.[28] Thus, we may parametrize the entropy

[28] As an example, for the case of a semi-classical gas of photons in D dimensions, the spectral energy density is given by Planck's law (reinstating constants) as

$$\rho(\nu) = 2 \left(\frac{\sqrt{\pi}}{c} \right)^D \frac{d-1}{\Gamma(D/2)} \frac{h\nu^D}{e^{h\nu/k_B T} - 1},$$

which yields an internal energy

$$E(T) = V_D \int_0^\infty d\nu \, \rho(\nu) = 2\,(D-1) \frac{\Gamma(D+1)}{\Gamma(D/2)} \operatorname{Li}_{D+1}(1) \left(\frac{\sqrt{\pi}}{hc} \right)^D k_B^{D+1} \, V_D T^{D+1},$$

where we have used the polylogarithm defined as $\operatorname{Li}_s(z) = \sum_{k=1}^\infty \frac{z^k}{k^s}$. Thus, from the thermodynamic relation (1.3.9), one can readily see that

$$S(T) = 2\, \frac{D^2-1}{D} \frac{\Gamma(D+1)}{\Gamma(D/2)} \operatorname{Li}_{D+1}(1) \left(\frac{\sqrt{\pi}}{hc} \right)^D k_B^{D+1} \, V_D T^D,$$

so indeed $S \sim V_D T^D$.

as

$$S_{\text{CFT}} = cT^{d-1}, \tag{3.1.1}$$

where c is a constant (which is dimensionless in our units) that can be regarded as a measure of the number of fields and particles associated with them.

Now let's consider the entropy in the bulk theory, this would scale as $S \sim V_d T^d$. Without loss of generality for this argument, it suffices to focus on a gas of gravitons (which are of course bosonic) that would be inherent to gravitational interactions in the AdS bulk, giving a lower bound on the entropy, as there could be other fields that have been excluded. We can further consider a finite region with $r \sim 1$, so that $S \sim T^d$, meaning that the temperature and entropy are of a similar order.[29] This must obviously be a lower bound of the gas of gravitons entropy, so we have

$$S_{\text{graviton gas}} > T^d. \tag{3.1.2}$$

Thus, we see that for large enough temperature when $T \sim b$, we have an apparent violation of the correspondence, as the entropy of the bulk would be greater than the one of the conformal boundary. But there is one key element that we are missing in this argument, which is that since the bulk must contain a theory of gravity, we must account for the formation of black holes. In particular, the metric for an AdS black hole takes the form

$$ds^2_{\text{AdS}_{d+1}} = L^2 \left[-\left(r^2 + 1 - \frac{r_+^d}{r^{d-2}}\right) d\tau^2 + \frac{dr^2}{r^2 + 1 - \frac{r_+^d}{r^{d-2}}} + r^2\, d\Omega^2_{d-1} \right], \tag{3.1.3}$$

where we see that r_+ is the horizon radius for $r^2 \gg 1$. It takes the form $r_+^d = gm$, where g is proportional to the Newton constant in units of the AdS radius of curvature and for $d \geq 3$ can be defined as

$$g = \frac{16\pi}{(d-1)\Omega_{d-1}} \frac{G_N^{(d+1)}}{L_{\text{AdS}}^{(d-1)}}, \tag{3.1.4}$$

so that m is the AdS version of the ADM mass of the black hole and we have included superscripts to emphasize the dimensions of the spaces that characterize these two scales.

This is the equivalent of the Schwarzchild solution in AdS, i.e. it describes a static, spherically symmetric spacetime with a singularity at the orgin that asymptotically approaches AdS[30] (instead of Minkowski).

The gas of gravitons has an effective size $r \sim T$ and mass $m \sim T^{d+1}$.[31] We then see that for large T (and thus r), then $r_+ = gm \sim gT^{d+1}$.

But now we see that at large enough temperatures, and in particular with $T > 1/g$, then the size of the system exceeds the Schwarzchild radius, meaning that the calculation can't be reliable in this limit as the gas of bosons would condense into a black hole. Therefore, in this high energy limit, we should not consider the entropy of a gas of

[29] Of course, if this was not the case, i.e. if the constant prefactor happened to be especially large or small, then we could simply adjust the radius of the region where we compute the entropy accordingly, so this is not really a problem.

[30] Indeed, we see that this reduces to the form of the AdS metric (2.1.8) as $r \to \infty$.

[31] This is since $m \sim E \sim V_d T^{d+1}$ (see Footnote 28).

bosons any more, but that of a black hole, which as we saw in Section 1.2 has the crucial property that it scales with the area, not volume.

AdS black holes follow a very similar relationship to the entropy formula (1.3.10), taking the form[32] $S \propto \frac{r_+^{d-1}}{g}$. They also have the equivalent of the Hawking temperature, which has the property that for large black holes $T_H \propto r_+$[33], which implies that

$$S_{\text{AdS bulk}} \propto \frac{T^{d-1}}{g}, \tag{3.1.5}$$

which is indeed consistent with the entropy of the CFT in (3.1.1) as required from AdS/CFT. Making these two entropies equal then yields

$$c \propto \frac{1}{g} \propto \frac{L_{\text{AdS}}^{(d-1)}}{G_N^{(d+1)}}. \tag{3.1.6}$$

Thus, we see that at high energies black holes dominate the AdS bulk and their degrees of freedom can be precisely mapped to those of a CFT in the boundary. This is one of the key realisations of AdS/CFT in relation to the information problem: black hole behaviour can be expressed as a manifestly unitary processes in a CFT.[34] It is also strongly suggestive of the fact that the entropy given by the area law truly describes internal degrees of freedom, as they are equivalent to those of an ordinary bosonic gas in a CFT with a well-defined partition function and associated thermal properties.

We also see from the above relationship that weakly-interacting gravitational theories in AdS units (i.e. $1/g \gg 1$), which is the classical limit of gravity with simpler interactions, are associated with a large number of fields in the CFT. In fact, a more detailed analysis involving Feynman diagrams and properties of operators associated with particle creation (like the stress-energy tensor and its traces, which produce gravitons in AdS) reveals that the classical limit of Einstein gravity in the AdS bulk is achieved when the CFT is strongly coupled.[35]

---

[32]Indeed, in the original equation $S \propto \frac{r_s^{3-1}}{G}$. This shouldn't be too surprising, we already saw in Section 1.4 that dividing a system into a partition that restricts exterior observers (like event horizons) produces area laws in the entropy and Bekenstein's argument need not to apply to ordinary Schwarzchild geometries only.

[33]The more general expression is

$$T_H = \frac{d - 2 + dr_+^2}{4\pi r_+},$$

which is quite different from the original Hawking temperature in (1.3.8) with

$$T_H \propto \frac{1}{r_s}.$$

Note also that $r_+ \not\propto gm$. Thus, owing to the significant differences between the two horizons, this is why we have (following other authors) used the notation r_+ instead to avoid potential confusions.

[34]AdS black holes also have another peculiarity: if sufficiently large, due to the lightlike finiteness of the boundary, the radiation gets reflected back and reabsorbed by the black hole, making an equilibrium state that doesn't evaporate. This is analogous to the state reached by pouring sufficient fluid into a closed container once it reaches thermal equilibrium. This may lead to another sense in which AdS spacetimes couldn't violate unitarity: it may be that external observers never measure a sufficiently large amount of radiation entropy (even classically), so that the Page time becomes infinite.

[35]This is a necessary but not sufficient condition.

3.2 The Ryu-Takayanagi Formula

We will now review the first successful instance of a full quantum treatment of the gravitational entropy that was proposed by Ryu and Takayanagi in 2006 by considering the simple case of AdS_3/CFT_2 [28, 29].

This is the spacetime that we previously discussed in Section 2.1 with metric (2.1.6) and whose Penrose diagram is shown in Fig. 2.1.2. The basic idea is that we can take a Cauchy slice of the cylinder (which has the topology of a disk) and then divide it into two regions and compute the von Neumann entropy of one region with respect to an observer in the other one. This is essentially the prescription we discussed in Section 1.4.

Motivated by AdS/CFT and the original BH entropy-area relation (1.3.10), this led to proposing the following definition for the entropy: consider a system A in a d-dimensional Cauchy slice of AdS_{d+1} whose entropy is defined by that of a $(d-1)$-dimensional boundary section ∂A of the $\mathbb{R} \times S^{d-1}$ conformal boundary. The proposal says that this entropy should (to leading order and in $\hbar = c = 1$ units) be given by

$$S_A = \frac{\text{Area of } \gamma_A}{4G_N^{(d+1)}}, \tag{3.2.1}$$

where γ_A is the static minimal surface of $d-1$ dimensions that traverses the Cauchy slice and whose boundary connects with ∂A.

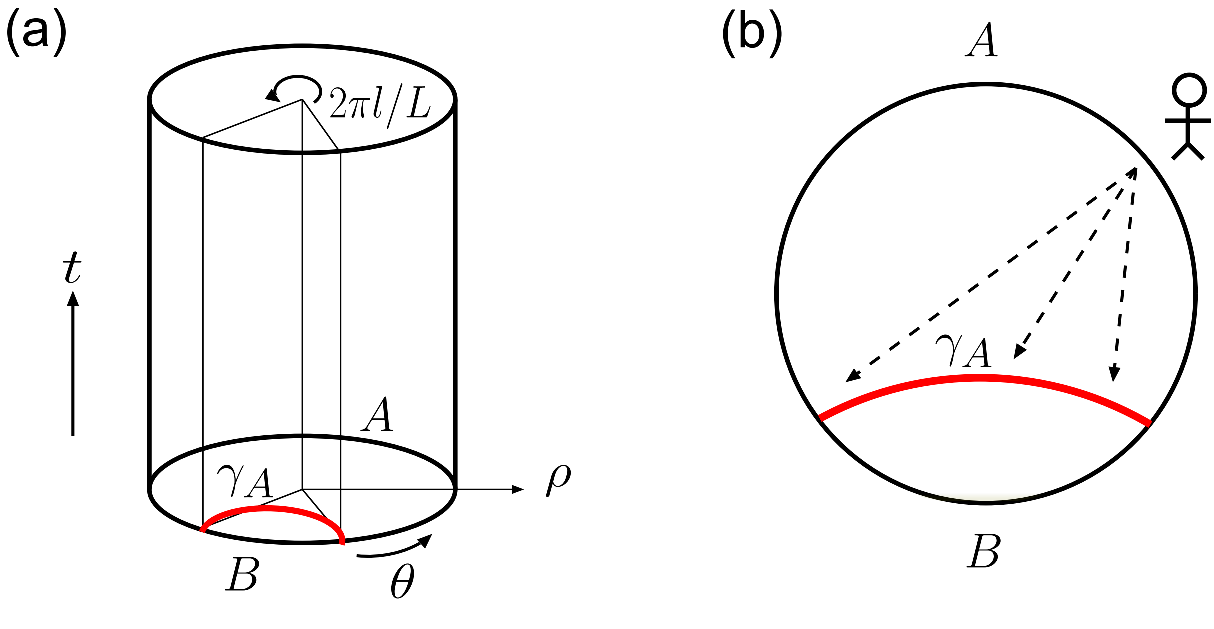


Image credit: [28]

Figure 3.2.1: The Ryu-Takayanagi prescription in AdS_3/CFT_2. (a) illustrates the AdS_3 cylinder where the entropy of a subsystem A is being computed. The minimal surface in this case is the geodesic connecting the two endpoints of the region, whose generalized area is simply its length. (b) illustrates how one can think of the entropy computed in B as that which would arise for an observer at the boundary of the complementary region A who is restricted from accessing B. Note that A and B are in the same Cauchy slice.

Although not explicit in this definition, this indeed looks like the entropy seen by an observer in the complementary region B who has no access to A (and vice versa). It also clearly satisfies the aforementioned property of $S_A = S_B$[36] (since both systems share the same minimal surface) and the subadditivity condition (1.4.6) (which is certainly true for CFTs describing pure states). Fig. 3.2.1 illustrates how this procedure is applied to the special case of AdS_3/CFT_2.

Eq. (3.2.1) is intended to provide a lower bound of the information loss to an observer who has been restricted to access the region were the entropy is computed, this is the reason why the minimal surface is taken. This is quite reminiscent of Bekenstein's holographic entropy bound which we discussed at the end of Section 1.3.

It is interesting to note that this bound is precisely saturated in AdS_3/CFT_2 (and plausibly in all higher dimensional cases too [28]). This is why this conjecture was motivated by AdS/CFT, and in the remainder of this section we shall discuss how this was done.

For CFT_2, the entropy of a given one-dimensional region in a Cauchy slice is known. At the region of criticality,[37] this is given by [28, 30, 31]

$$S_A = \frac{c}{3} \ln \left[\frac{L_{\text{tot}}}{\pi a} \sin \frac{\pi l}{L_{\text{tot}}} \right], \tag{3.2.2}$$

where l and L_{tot} are the lengths of the subsystem and total system $A \cup B$ respectively and periodic boundary conditions have been applied (so that the conformal boundary in the Cauchy slice has the S^1 topology of a circle). a is the UV cut-off and can be interpreted as the spacing of the lattice. c is the central charge that characterizes the CFT.

Outside criticality, the expression changes to [28, 30, 32]

$$S_A = \frac{c}{6} \mathcal{A} \ln \frac{\xi}{a}, \tag{3.2.3}$$

where ξ is the correlation length (a measure of spatial statistical dependence between the constituent elements of the system) and $\mathcal{A}$ the number of boundary points (2 in our case).

From AdS/CFT, gravity in the AdS_3 bulk with radius of curvature $L^{(1)}_{\text{AdS}}$ should be dual to a CFT_2 (i.e., in $1+1$ dimensions) with central charge [28, 33]

$$c = \frac{3 L^{(1)}_{\text{AdS}}}{2 G^{(3)}_N}. \tag{3.2.4}$$

Note that by interpreting the central charge as a measure of the effective number of fields in the CFT, this equation is indeed consistent with what is expected from Eq. (3.1.6).

[36] This is true for $A \cup B$ being a pure state, if the system has an effective temperature $T > 0$, this would imply that it is in a generally mixed state and the equality wouldn't hold.

[37] CFTs turn out to be particularly useful for studying critical phenomena. For example, the Ising model, which describes how atomic spins in lattices align with each other to produce magnetism, has a second order phase transition from a magnetic (ferromagnetic) to a non-magnetic (anti-ferromagnetic) phase. At this critical point the correlation length diverges and the system becomes scale invariant, which is of course a conformal symmetry.

As discussed previously, the AdS_3 metric (2.1.6) has an IR divergence as $\rho \to \infty$. To avoid this issue, we need to restrict the space to a finite region defined by $\rho \leq \rho_0$. This upper limit is dual to a UV cut-off in the boundary CFT [28, 34].[38] For the system discussed above, this cut-off satisfies the relation [28]

$$e^{\rho_0} \sim \frac{L_{\text{tot}}}{a}. \tag{3.2.5}$$

We are now left with the geometry of a finite cylinder of radius $L^{(1)}_{\text{AdS}}\rho_0$. As seen in Fig. 3.2.1,[39] from elementary geometry we can parametrize the length along the circular arc of total length l in A with the angle interval $0 < \theta < 2\pi l/L_{\text{tot}}$. The minimal surface is then given by the length of the geodesic connecting the endpoints at the boundary which are at $\theta = 0$ and $\theta = 2\pi l/L_{\text{tot}}$.

Note that the length along the circumference of this circle is not simply $L^{(1)}_{\text{AdS}}\rho_0\,\theta$. In particular, from the metric in global coordinates (2.1.6), we see that an arc along this circle has length

$$\ell = L^{(1)}_{\text{AdS}} \sinh\rho_0 \int_0^\theta d\theta' = L^{(1)}_{\text{AdS}} \sinh\rho_0\, \theta. \tag{3.2.6}$$

This motivates the definition of an effective (dimensionless) radial coordinate $r = \sinh\rho$, which is indeed the one used in the metric (2.1.8). This discrepancy with Euclidean geometry arises due the fact that we are considering the equivalent of a circle in a conformally compactified space, not the usual Cartesian plane.[40]

We can use this result for the circumference to get a rough upper bound of the entropy of this system. In particular, we can compute a one-dimensional area analog of the BH entropy (1.3.10) as

$$S_{\max} \sim \frac{\text{Area of boundary}}{4G_N^{(3)}} = \frac{2\pi L^{(1)}_{\text{AdS}}\sinh\rho_0}{4G_N^{(3)}} \approx \frac{\pi c}{6} e^{\rho_0} \sim \frac{\pi c}{6}\frac{L_{\text{tot}}}{a}, \tag{3.2.7}$$

where we have used the relations (3.2.4) and (3.2.5), and also taken $\rho_0 \gg 1$. We see how the UV cut-off a behaves like a fundamental unit of entropy similar to the Planck area (up to a factorization of constants). We know this should be an upper bound because the length of a geodesic cannot be greater than that of the circumference of the circle.[41]

Now, with that in mind, let's proceed to calculate the entropy by properly following the Ryu-Takayanagi prescription. To do this, we must first obtain the expression for a

[38] This is another interesting entry of the AdS/CFT dictionary: the IR physics in the bulk theory side is dual to the UV physics at the boundary. The study of how the dual divergences of these two regimes arise constitutes the field known as holographic renormalization. For more information, see [35].

[39] As a minor point on notation, note that in the figure, which is the original one found in [28], the authors use L in place of L_{tot} for the total circumference of the cylinder (not to be confused with the AdS radius $L^{(1)}_{\text{AdS}}$, which was simply denoted as L in previous sections) and t in place of the time dimension τ found in the metric (2.1.6).

[40] This illustrates how conformal compactification preserves angles, but not lengths. In particular, we see that in the Penrose diagram the radial length is indeed perpendicular to the circumference of the circle, but its arclength is no longer proportional to it.

[41] If it were, this would be a contradiction with the definition of a geodesic, since any point in the boundary can be connected through a circular path that is lesser or equal to the circumference.

geodesic in AdS_3. In this case, it is most useful to consider Poincaré coordinates (2.1.11), wich yield the metric (2.1.12).

By the usual straightforward computation (i.e. computing Christoffel symbols etc.), one readily obtains the affinely-parametrized geodesic equations

$$\begin{aligned} \frac{\mathrm{d}^2 x}{\mathrm{d}\lambda^2} - \frac{2}{z}\left(\frac{\mathrm{d}x}{\mathrm{d}\lambda}\right)\left(\frac{\mathrm{d}z}{\mathrm{d}\lambda}\right) &= 0, \\ \frac{\mathrm{d}^2 z}{\mathrm{d}\lambda^2} + \frac{1}{z}\left[\left(\frac{\mathrm{d}x}{\mathrm{d}\lambda}\right)^2 - \left(\frac{\mathrm{d}z}{\mathrm{d}\lambda}\right)^2\right] &= 0. \end{aligned} \tag{3.2.8}$$

It can be readily checked that solutions to these equations can be parametrized as

$$\begin{aligned} x(\alpha) &= \frac{l}{2}\cos\alpha, \\ z(\alpha) &= \frac{l}{2}\sin\alpha, \end{aligned} \tag{3.2.9}$$

where

$$\alpha(s) = 2\tan^{-1}\left[e^{s/L^{(1)}_{\mathrm{AdS}}}\right], \tag{3.2.10}$$

and s is the affine parameter describing the length along the geodesic.

The constant-time Poincaré metric is given by

$$ds^2 = \left(\frac{L^{(1)}_{\mathrm{AdS}}}{z}\right)^2 [dx^2 + dz^2]. \tag{3.2.11}$$

We then readily find from the above parametrization (3.2.10) that this has a line element

$$ds = \frac{L^{(1)}_{\mathrm{AdS}}}{\sin\alpha}\, d\alpha. \tag{3.2.12}$$

By integrating both sides we get

$$s = L^{(1)}_{\mathrm{AdS}} \ln[\tan(\alpha/2)] + \text{constant}, \tag{3.2.13}$$

which is indeed consistent with the affine parametrization (3.2.10).

As we approach the boundary as $z \to 0$ and $\rho \to \infty$, this geodesic describes a circle of radius $L^{(1)}_{\mathrm{AdS}} l/2$ connecting two nearby points. This is illustrated in Fig. 3.2.2. The cut-off ρ_0 leads to a UV cut-off for the z coordinate z_{UV} and an angular one ϵ so that $\epsilon < \alpha < \pi - \epsilon$. In this boundary limit, these cut-offs are geometrically related by the simple relation $z_{\mathrm{UV}} \approx l\epsilon/2$. Also in this limit, z_{UV} effectively becomes the boundary CFT UV cut-off a.[42]

[42]This can be seen from the fact that since, using (2.1.11), $x/z = \sin\theta\tan\chi = \sin\theta\sinh\rho$, then near the boundary $e^\rho \sim x/z$, which can be directly compared with the relation (3.2.5) for $\rho = \rho_0$ and $x = L_{\mathrm{tot}}$ to give $z_{\mathrm{UV}} \sim a$. Note that there is an ambiguity in the proportionality factor, but this turns out to be irrelevant for computing the entropy.

We can now compute the regulated length of the geodesic as

$$L_{\gamma_A} = \int_{s(\alpha=\epsilon)}^{s(\alpha=\pi-\epsilon)} ds = 2L_{\mathrm{AdS}}^{(1)} \int_{\epsilon}^{\pi/2} \frac{d\alpha}{\sin\alpha} = -2L_{\mathrm{AdS}}^{(1)} \ln[\tan(\epsilon/2)] \approx 2L_{\mathrm{AdS}}^{(1)} \ln\frac{l}{a}, \tag{3.2.14}$$

where for convenience we have exploited the fact the geodesic is an even function in the x-z plane of the setup as shown in Fig. 3.2.2 and we have also taken the limit $l \gg a$.

From the proposal (3.2.1) and the dual relation (3.2.4), this now yields the entropy

$$S_A = \frac{L_{\mathrm{AdS}}^{(1)}}{2G_N^{(3)}} \ln\frac{l}{a} = \frac{c}{3} \ln\frac{l}{a}, \tag{3.2.15}$$

which indeed agrees with the original CFT expression (3.2.2) for $l \ll L_{\mathrm{tot}}$.

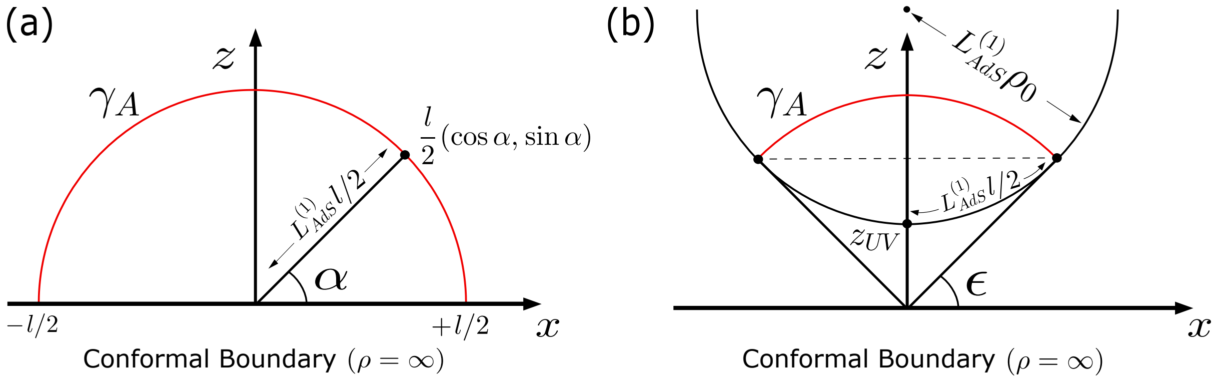


Image credit: Own work.

Figure 3.2.2: The setup for computing geodesics in Poincaré coordinates. (a) shows the cut-off limit of $\rho_0 \to \infty$ and $\epsilon, z_{\mathrm{UV}} \to 0$. For closely separated points by a (dimensionless) distance l in the x coordinate at the boundary, with $a \ll l \ll L_{\mathrm{tot}}$, the geodesic between the two is given by the semicircle of (dimensionful) radius $L_{\mathrm{AdS}}^{(1)} l/2$ connecting them. (b) shows the geometry associated with these three cut-offs. The geodesic between points in the circle has a geometry equivalent to that of a circular arc. This statement can be generalized to arbitrary points near the region, where the geodesic is given by constructing a circular arc connecting them which perpendicularly intersects two points at the boundary. This is the geometry of the Poincaré disk.

A more general expression can be obtained for the geodesic in AdS_{d+1} by considering the intersection between the surface defined by this space and a two-dimensional hyperplane in the ambient space $\mathbb{R}^{2,d}$. The geodesic is then given by the intersection where the normal vectors to its points are also in the hyperplane. For the special case of AdS_3, this yields the relation [28, 29]

$$\cosh\frac{L_{\gamma_A}}{L_{\mathrm{AdS}}^{(1)}} = 1 + 2\sinh^2\rho_0 \sin^2\frac{\pi l}{L_{\mathrm{tot}}}. \tag{3.2.16}$$

By taking $e^{\rho_0} \gg 1$ and $L_{\gamma_A} \gg L^{(1)}_{\text{AdS}}$ we obtain

$$L_{\gamma_A} \approx L^{(1)}_{\text{AdS}} \ln\left[e^{2\rho_0} \sin^2 \frac{\pi l}{L_{\text{tot}}}\right]. \tag{3.2.17}$$

Now applying again the proposal (3.2.1), the dual relation (3.2.4) and also (3.2.5), we get

$$S_A \approx \frac{L^{(1)}_{\text{AdS}}}{2G^{(3)}_N} \ln\left[e^{\rho_0} \sin \frac{\pi l}{L_{\text{tot}}}\right] = \frac{c}{3} \ln\left[\frac{L_{\text{tot}}}{\pi a} \sin \frac{\pi l}{L_{\text{tot}}}\right] + \text{constant}, \tag{3.2.18}$$

where the constant term needs to be added to account for the ambiguity in the factor of proportionality in (3.2.5), which can however be ignored for our purposes.

We see that this result precisely agrees with the original CFT entropy (3.2.2), so that, once more, we have seen an explicit realisation of AdS/CFT and the success of this proposal for computing the entropy.

Let's consider what happens for the minimal arc length $l = a$. Since $L_{\text{tot}} \gg a$, we clearly get $S_A = 0$, as all the terms inside the logarithm cancel. We can similarly consider the complement region with $l = L_{\text{tot}} - a$, which in this case expanding about π gives $\sin \frac{\pi l}{L_{\text{tot}}} \approx \pi(1 - \frac{L_{\text{tot}}-a}{L_{\text{tot}}}) = \frac{\pi a}{L_{\text{tot}}}$, so once again the entropy vanishes. This is consistent with the expected properties of pure states divided into bipartite systems. Geometrically, one can think of this as a consequence of the geodesic shrinking to zero size as the endpoints of the boundaries get closer together.

It is also easy to see that the entropy is maximized when at $l = L_{\text{tot}}/2$ (i.e. when the sine is 1). This yields the entropy $S_{\text{max}} = \frac{c}{3} \ln \frac{L_{\text{tot}}}{\pi a}$, which clearly satisfies the initial bound (4.1.1), since $L_{\text{tot}} \gg a$ and it is a standard property of logarithms that $x \gg \ln x$ for $x \gg 1$.

In their main paper [28], Ryu and Takayanagi also showed further examples of applying this prescription to cases of thermal systems with non-zero temperature, higher dimensional cases of AdS/CFT, and cases involving multiple (non-bipartite) systems. However, conceptually speaking, the approaches are essentially the same. The case we have reviewed is the most instructive and the one that they originally used to formulate their prescription.

3.3 HRT Proposal and General Covariance of the Entropy

The AdS_3 spacetime studied by Ryu and Takayanagi (RT), while useful for motivating a framework for computing the entropy, suffers from a notable limitation: it is symmetric under time translations, i.e. there is a preferred Cauchy slice whose geodesics have no time dependence (and thus the entropy derived from these). This should be quite clear, as at no point in the previous section did we need to specify the time, it sufficed to know that it was set constant.

However, if one wishes to reproduce Page curves, time dependence is crucial. We know that the entropy of black hole evaporation as seen by an outside observer should be dynamic. It is also clear that generic spacetimes will not have this time translation symmetry and could yield multiple Cauchy slices with differing values for the entropy. What one needs is a covariant generalization of the RT formula. This was provided by Hubeny, Rangamani and Takayanagi (HRT) [36].

Interestingly, the approach used by HRT is based on a covariant generalization of the Bekenstein bound discussed in Section 1.3 that was found by Bousso [37, 38].[43] In particular, Bekenstein's original proposal that the entropy of some arbitrary region in space enclosed by a sphere of area A will always be equal or lesser than $A/4G_N$, is routinely violated in cosmology and it is not a covariantly well-posed statement, as it implies some particular preferred reference frame. For example, one could imagine boosting along a coordinate axis so that the sphere contracts and then"stacking" an arbitrarily large number of such spheres inside a finite surface, violating the bound.

Such a covariant generalization is constructed by imagining each point in the surface following a null geodesic. Such geodesics have four possibilities: ingoing/outgoing future-directed and ingoing/outgoing past-directed. However, these must behave in such a way so as to not increase the size of the resultant surface over time. The most instructive possibility is to imagine an initial sphere whose radius contracts at the speed of light. The Bousso bound then says that the entropy of such a shrinking region is bounded by the original Bekenstein bound of the initial surface $A/4G_N$.[44] This case is illustrated in Fig. 3.3.1. The congruence of null geodesics that bounds the area of the region in spacetime is known as a light-sheet.

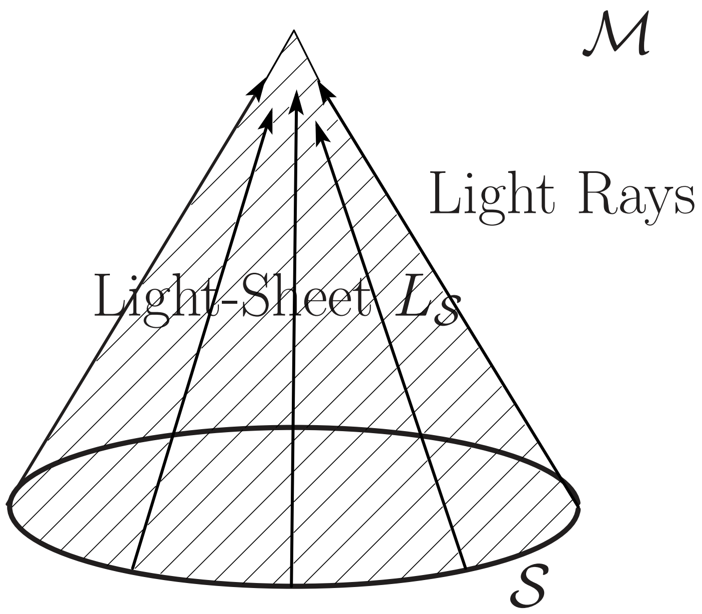


Image credit: [36]

Figure 3.3.1: Spacetime diagram of the conical light-sheet $L_\mathcal{S}$ bounding space-like regions in a flat space $\mathcal{M}$. The regions enclosed by Cauchy slices in $L_\mathcal{S}$ have an entropy bounded by the area of the surface $\mathcal{S}$ at the base. Beyond the focal point at the apex, the surface would start to increase, meaning that this would no longer be part of the bound.

[43] This has also been used in the context of extending the Generalized Second Law, see [39, 40].

[44] One way to think about this is that, by contracting the region in such a way, one is left with a "locally safe" region where no influx of entropy from the outside (e.g. due to a flow of particles in mixed states) could access such a region without violating causality.

Also motivated by AdS/CFT, the way HRT used the Bousso bound to generalize the RT prescription was as follows [36]: Consider an AdS_{d+1} spacetime $\mathcal{M}$ with the usual $\mathbb{R}\times S^{d-1}$ boundary $\partial\mathcal{M}$. Now divide the region in a Cauchy slice into a bipartite system with boundary portions $\mathcal{A}_t$ and $\mathcal{B}_t$, with a $(d-2)$-dimensional partition boundary $\partial\mathcal{A}_t$ splitting them. The t subscripts are placed to emphasize a potential time-dependence.

To proceed, for simplicity we use Poincaré coordinates which yield the manifestly conformally-flat metric (like the one in (2.1.12)) on the boundary $\partial\mathcal{M}$. One now readily constructs two $(d-2)$-dimensional light-sheets ∂L_t^+ and ∂L_t^- (one future-directed and another one past-directed) for $\partial\mathcal{A}_t$ in $\partial\mathcal{M}$. This then allows forming a $(d-1)$-dimensional congruence of $\partial L_t^\pm$, $L_t^\pm$, which defines two light-sheets (respectively) in $\mathcal{M}$ that extend into the bulk with respect to a space-like $(d-1)$-dimensional surface $\mathcal{Y}_t = L_t^+\cap L_t^-$. This setup is summarized in Fig. 3.3.2.

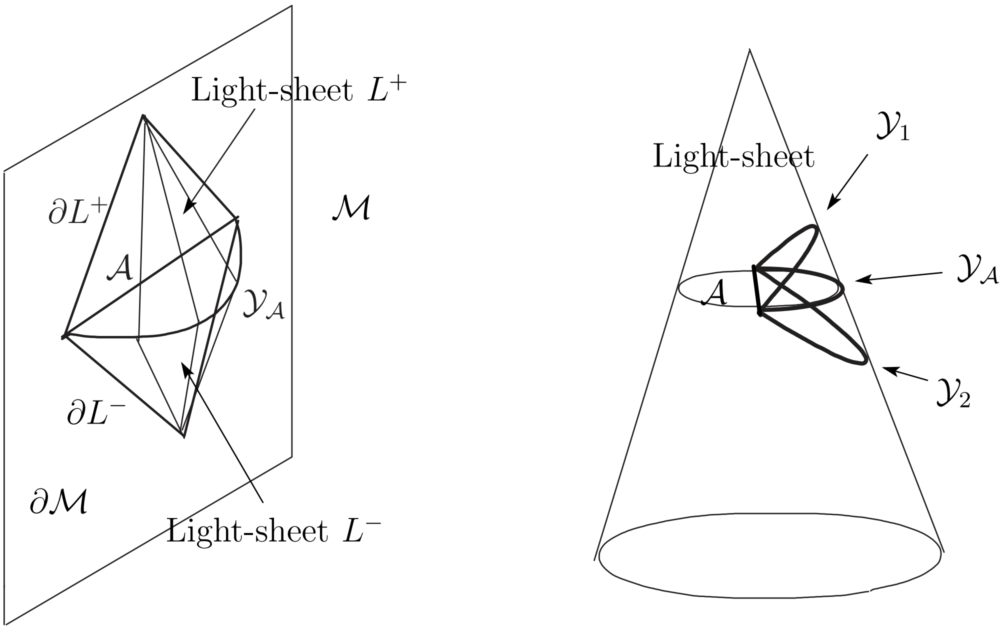


Image credit: [36]

Figure 3.3.2: The light-sheet construction for the HRT proposal we have described applied to the familiar case of $\mathrm{AdS}_3/\mathrm{CFT}_2$. For this special case the HRT proposal reduces to computing the entropy via the length of the geodesic $\mathcal{Y}_\mathcal{A}$, as per the RT prescription in the previous section. $\mathcal{Y}_{1,2}$ denote arbitrary, non-geodesic paths within the light-sheet. Note that t subscripts have been omitted for notational simplicity.

With the above in mind, the HRT proposal then states that the generally time-dependent entropy for the subsystem $\mathcal{A}_t$ is given by the minimal surface of $\mathcal{Y}_t$ as

$$S_{\mathcal{A}_t} = \frac{\min_{\mathcal{Y}}[\mathrm{Area}(\mathcal{Y}_t)]}{4G_N^{(d+1)}}, \tag{3.3.1}$$

where, since the choice of the bulk light-sheets $L_t^\pm$ is not unique in general, one needs to find the global minimum of the area from the boundary $\mathcal{Y}_t$ that would correspond to iterating over all choices of $L_t^{\pm}$ with $\partial L_t^{\pm}$ fixed.

The relation (3.3.1) can be written in a more suggestive form that allows for direct comparison with the RT result (3.2.1) as

$$S_A(t) = \frac{\text{Area of } \gamma_A(t)}{4G_N^{(d+1)}}, \tag{3.3.2}$$

where $\gamma_A(t)$ can be defined as the minimal surface over the whole (Lorentzian) spacetime satisfying the boundary condition $\partial\gamma_A(t) = \partial A_t$ [41].

It should be noted that a general definition of an area involving a Lorentzian signature in the metric is a non-trivial problem. This was not an issue with the original RT proposal, since the time dimension effectively vanished for Cauchy slices and one was left with a Euclidean metric, but it is a key consideration for allowing a covariant generalization. For example, if one is not careful about this, an arbitrarily large time-like contribution could make the area of the surface vanishing,[45] which would preclude any meaningful notion of a minimal surface.

This led to an important addendum to the HRT proposal, or what its authors called the "proposal II" [36], which is based on formulating a variational principle for finding extremal surfaces. In particular, this is done by a variation of the area functional due to a deformation caused by the expansion of a surface along the null direction, as originally described in the Bousso bound with light-sheets. By considering a spacetime $\mathcal{M}$ and defining the embedding map $X^\mu(\xi^\alpha)$ from coordinates ξ^α in some codimension-2 (two dimensions less than $\mathcal{M}$) surface $\mathcal{S}$ to $\mathcal{M}$ as the ambient spacetime, then the resulting variation is given by

$$\delta\text{Area} \propto \int_{\mathcal{S}} (\theta_+ \, N_+^\mu \, \delta X_\mu + \theta_- \, N_-^\mu \, \delta X_\mu), \tag{3.3.3}$$

for some positive constant of proportionality, where $N_\pm^\mu$ are two linearly-independent vectors orthogonal to surface $\mathcal{S}$ under consideration and are normalized so as to satisfy

$$g_{\mu\nu} \, N_+^\mu \, N_-^\nu = -1. \tag{3.3.4}$$

$\delta X_\pm^\mu$ represent the respective infinitesimal surface deformations whose contribution will be in the directions orthogonal to $\mathcal{S}$. $\theta_\pm$ are two parameters known as the null expansions and are given by the trace of the null extrinsic curvature $(\chi_\pm)_{\mu\nu}$, which can be defined in terms of the induced metric $h_{\mu\nu}$ and the surface normal vectors as

$$(\chi_\pm)_{\mu\nu} = h^\rho{}_\mu \, h^\lambda{}_\nu \, \partial_\rho (N_\pm)_\lambda, \tag{3.3.5}$$

so that

$$\theta_\pm = (\chi_\pm)^\mu{}_\mu. \tag{3.3.6}$$

As their name suggests, null expansions quantify the rate of change in the area due to infinitesimal displacements in the directions orthogonal to the surface (i.e. along the null

[45] For example, a null path connecting two boundary points in AdS_3 would have a vanishing "length".

vectors $N^\mu_\pm$) of the points that lie in it. Specifically, $\theta_\pm < 0$ makes the variation in the area functional (3.3.3) negative (and conversely positive for positive $\theta_\pm$), which gives a precise interpretation of the key requirement formulated in the original Bousso bound that surface expansions, as described by light-sheets, must lead to non-positive changes in the area. This also implies that surfaces whose null expansions vanish ($\theta_\pm = 0$) will be saddle points in the area functional from (3.3.3), thus allowing for a suitable general framework to covariantly compute extremal surfaces.[46]

3.4 FLM/EW Proposals and Quantum Extremal Islands

So far, we have given a classical treatment of the entropy in the sense that the methods that have been discussed are purely geometrical in nature. However, Hawking radiation has a manifestly non-classical origin: quantum instabilities in the QFT vacuum. This means that the spacetimes we should be interested in are those which are coupled to some non-trivial, dynamic QFT, leading to foreseeable quantum corrections in the entropy.

Although it was once thought that a full quantum treatment of gravity would be needed for this, recent lines of research suggest that, for the most part, a semi-classical treatment involving perturbative corrections to the entropy suffices. This is quite remarkable, as it means that one can gain crucial insights on the quantum evolution of gravitational systems from a relatively simple standpoint. As we will see, in our context it establishes a prescription for computing the entropy of Hawking radiation, from which one obtains Page curves that serve as independent evidence of unitary evolution as a resolution to the information problem [42, 43, 44].

Such a proposal was obtained by Faulkner, Lewkowycz, and Maldacena (FLM) [45] and later further generalized by Engelhardt and Wall (EW) [46]. The basic idea is that one can obtain a perturbative expansion in $\hbar$ for the entropy. The classical area-entropy relation is then given as the leading $\mathcal{O}(\hbar^{-1})$ term, with a sub-leading $\mathcal{O}(\hbar^0)$ correction that is due to the von Neumann entanglement entropy of the quantum fields coupled to the (otherwise classical) spacetime. This can be expressed as [45, 46][47]

$$S_R = \frac{\langle \text{Area of } \partial X \rangle}{4G_N\hbar} + S_{\text{ent}}(R \cup X) + \text{counterterms} + \mathcal{O}(\hbar) = S_{\text{gen}}(X), \tag{3.4.1}$$

where X is the bulk region with surface ∂X under consideration that is homologous to a region of interest R in the CFT (à la the RT/HRT prescriptions).[48] S_{ent} is the (first order) bulk entanglement entropy due to ∂X acting as a partition. Note that we have reinstated $\hbar$ in the entropy formula to carry the expansion in the perturbation. Counterterms need to be added to deal with the divergences that arise. The leading divergence is proportional to the area [17, 46], we already saw a hint of this in Eq. (1.4.13), where it was shown on simple generic grounds that dividing a quantum system into a partition leads to an area law for the entropy proportional to the square of an (otherwise divergent) UV cut-off

[46] As usual, if the extrema obtained are not unique, then one should take the global minimum of these.

[47] This is taken to be in the absence of graviton fluctuations, which would otherwise lead to $\mathcal{O}(\hbar^{1/2})$ terms in the metric expansion [46].

[48] Note also that the area is written as an expectation, this is because its definition involves the smearing of curvature operators (which aren't well-defined at points) [47, 48, 49].

(which would correspond to an inverse lattice spacing in the CFT). Interestingly, the counterterms can be interpreted as a renormalization of the gravitational constant [50].

The main difference between the FLM and EW prescriptions, is that FLM extremized the area purely on geometrical grounds (as in RT/HRT) and then added the quantum correction term in Eq. (3.4.1), while EW proposed that the area should be extremized with respect to the whole entropy functional (including quantum corrections). This leads to what are known as Quantum Extremal Surfaces (QES) which are defined as those whose area extremize (3.4.1) as a whole. EW also extended the proposal to non-static spacetimes, as opposed to FLM originally.

FLM proved their proposal at order $\mathcal{O}(\hbar^0)$ by considering quantum corrections to the Euclidean gravitational path integral [51]. This means that the FLM and EW proposals are equivalent at this order, which was shown in [46], although there are non-trivial checks that suggest that EW is the more valid approach at higher orders, where the results of the proposals differ [46].

In their original derivation, FLM considered the canonical case of $1+1$ dimensional dilaton gravity, which has an action

$$I_{\text{grav}}[g_{ij}^{(2)},\phi] = \int d^2y\sqrt{-g}\left(\frac{1}{16\pi G_N^{(2)}}\,\phi R^{(2)} + V(\phi)\right). \tag{3.4.2}$$

We see that the Einstein-Hilbert term (which is proportional to the Ricci scalar $R^{(2)}$) can be trivially removed in this case by a shift of the scalar “dilaton field” ϕ. Under the substitution $\phi/G_N^{(2)} \equiv 1/G$, it is manifest that the action (3.4.2) can be interpreted as a modification of Einstein gravity with a varying gravitational constant, this was originally proposed in the context of Brans-Dicke theory [52].

The essential idea behind their proof is that one can then consider the path integral defining the effective partition function via the Euclidean version of the gravitational action as $\ln Z = -I_{\text{E, grav}}$ (which can be extended to include potential quantum corrections), from which one obtains the entropy via the analytically-continued Rényi prescription [51]

$$S = -n\partial_n[\ln Z(n) - n\ln Z(1)]|_{n=1}. \tag{3.4.3}$$

The motivation for this approach as a means to obtaining the entropy is as follows: The generically defined Rényi entropy for $n \geq 0$ is given by

$$S_n \equiv \frac{1}{1-n}\ln \mathrm{Tr}\rho^n. \tag{3.4.4}$$

This reduces to the von Neumann entropy (1.4.2) in the $n \to 1$ limit[49] and can be obtained

[49]To see this explicitly, recall that the eigenvalues of ρ are the Shannon probabilities p_i, so we see that

$$S_n = \frac{1}{1-n}\ln\sum_i p_i^n.$$

In the $n \to 1$ limit we have a $0/0$ type convergence that can be dealt with l'Hôpital's rule, which gives

$$\lim_{n\to 1} S_n = -\lim_{n\to 1}\frac{1}{\sum_i p_i^n}\sum_j p_j^n \ln p_j = -\sum_i p_i \ln p_i,$$

which is indeed the von Neumann entropy as found in (1.4.11).

for the special case of integer n greater than 1 by [53]

$$S_n = \frac{1}{1-n} \ln \frac{Z(n)}{Z^n(1)}, \tag{3.4.5}$$

where, for CFTs with a gravity dual, $Z(n)$ can be defined as the partition function of the theory living in what is known as the "n-fold branched cover" manifold. Roughly speaking, this corresponds to glueing n copies of the system to the boundary and is known as the "replica trick."

One can then analytically continue to non-integer values of n and define the closely associated quantity [53]

$$\widetilde{S}_n \equiv -n^2 \partial_n \left(\frac{1}{n} \ln \mathrm{Tr} \rho^n \right), \tag{3.4.6}$$

which by comparing Eqs. (3.4.4) and (3.4.5), exactly matches with the entropy in (3.4.3).

This leads to a very important realisation: one doesn't need to know the density matrix, i.e. the full quantum state of the system, to know its entropy. This means that the semi-classical description is not merely an approximation, it (in principle) has access to key thermodynamic aspects of the system. It strongly suggests that Nature has been kind enough to make these properties accessible without the need of an elusive theory of quantum gravity.

Motivated by this replica trick in two-dimensional dilaton gravity, this has led to a recent conjecture extending the notion of the generalized entropy [42]. This involves what are known as "islands", which are disconnected regions that extremize the entropy functional. In principle, there is no cap on how many such multiple surfaces one should include, so long as they are extremal. When considering such surfaces, there is a natural trade-off that one needs to consider: islands arise when, by including an external region to the usual surface, one is able to overcome the cost of the area term in the generalized entropy functional with the entanglement entropy. This happens because, by considering pairs of entangled radiation quanta, we can purify them and eliminate the resulting von Neumann entropy that arises from the inaccessibility to those "partner" exterior modes.

This new proposal also makes use of the maxi-minimization procedure that was originally developed by Wall [54]. The idea is that if one is to consider the entropy of some region R as in the previous prescriptions, then one should consider all possible Cauchy slices containing R, compute the minimal extremal surface via the usual methods, and then take the global maximum over such family of possible Cauchy-slice minima, hence the maxi-minimization.

Wall showed that this prescription satisfies several non-trivial properties expected from the entropy and its equivalence with the HRT prescription in classical settings [54].[50] It has also been recently given a more general interpretation in the context of "quantum maxi-min surfaces" [49], which is the idea that this procedure should be applied to the

[50] This equivalence with HRT, and in particular the "proposal II" we discussed, is not *a priori* obvious, which has led to some difficulties in proving theorems that entropies should satisfy, like for example, properties associated to subadditivity and "null energy type" conditions that are required for extensions of the generalized second law [49].

whole entropy functional, instead of the area alone classically. This is much like the distinction between the EW and FLM prescriptions.

The conjectured proposal, for some QFT region R, in terms of the maxi-min construction can then be expressed as [42, 49]

$$\widetilde{S_R} = \max_C \left\{ \min_{\mathcal{I}_g \in C} \left[\frac{\langle \text{Area of } \partial\mathcal{I}_g \rangle}{4G_N \hbar} + S_{\text{eff}}(R \cup \mathcal{I}_g) \right] \right\} = \max_C \left\{ \min_{\mathcal{I}_g \in C} \left[S_{\text{gen}}(\mathcal{I}_g) \right] \right\}, \quad (3.4.7)$$

where we have been careful to use a notation that allows for a direct comparison with the original FLM/EW proposal (3.4.1) in the cases where the QFT has a gravitational dual in the AdS/CFT sense.[51] C denotes the set of all choices for a Cauchy slice enclosing R. $\mathcal{I}_g$ stands for the island regions, the g is to emphasize the gravitational nature of the bulk theory in which they would be found in.[52] The islands obtained via this maxi-minimization procedure are known as "quantum extremal islands." S_{eff} illustrates the fact that we are dealing with the entropy of an effective field theory: It does not come from a full theory of quantum gravity and in practice is carried at some finite order in the perturbative expansion, in FLM/EW terms, it can be interpreted as "everything in (3.4.1) except the area term." Note also that (again, in the FLM/EW context) $\widetilde{S_R} = \max_C\{\min_{\mathcal{I}_g \in C}[S_R]\}$, so the former is the maxi-minimization the latter, meaning that the two entropies are not necessarily the same.

While the $\widetilde{S_R}$ notation has a formal difference as far as the semantics of (3.4.7) vs (3.4.1) are concerned, there is a deeper significance. There is a sense in which this entropy is meant to be *the* entropy, not only does it encapsulate all the previous proposals that we have discussed (which are all special cases of it), it is meant to encode in some non-arbitrary, non-generic sense the physics of the system. It is the kind of entropy that should be plotted in time to yield a meaningful Page curve of the system (in fact, this is what its authors originally used it for [42]).

This is not to say that the proposal is necessarily true, in fact, it has yet to pass some pending non-trivial checks [49]. The point is that, at least conceptually, this is the kind of entropy that should be meaningful to discuss when it comes to solutions of the information problem, even though the precise way it should be defined is yet to become a settled issue. The journey towards such a final proposal dates back to Bekenstein and Hawking in the 70s [6, 8], although the conundrum of the relation between entropy and information is at least as old as Maxwell's demon.

Fig. 3.4.1 illustrates how the proposal is applied in the original case of two-dimensional gravity coupled to a CFT_2 that motivated it. This example was studied in the classical gravity / strongly coupled CFT limit, i.e. a codimension 1 QFT with weak gravity in a higher dimensional bulk. An intriguing feature of such models is their close similarity with braneworlds as in the Randall-Sundrum model [55]. These models describe how one can have geometries effectively confined to lower-dimensional subspaces so that forces

[51] This is to make terms like "boundary" and "bulk" meaningful, although it is conceivable that this proposal can be interpreted so as to not be limited to theories with holographic duals.

[52] Strictly speaking, to allow for full consistency with cases involving a single region X as in (3.4.1), we could define X to be simply the "trivial island," i.e. the special case where there are no multiple disconnected regions.

lead to dominant interactions that are restricted to these.[53]

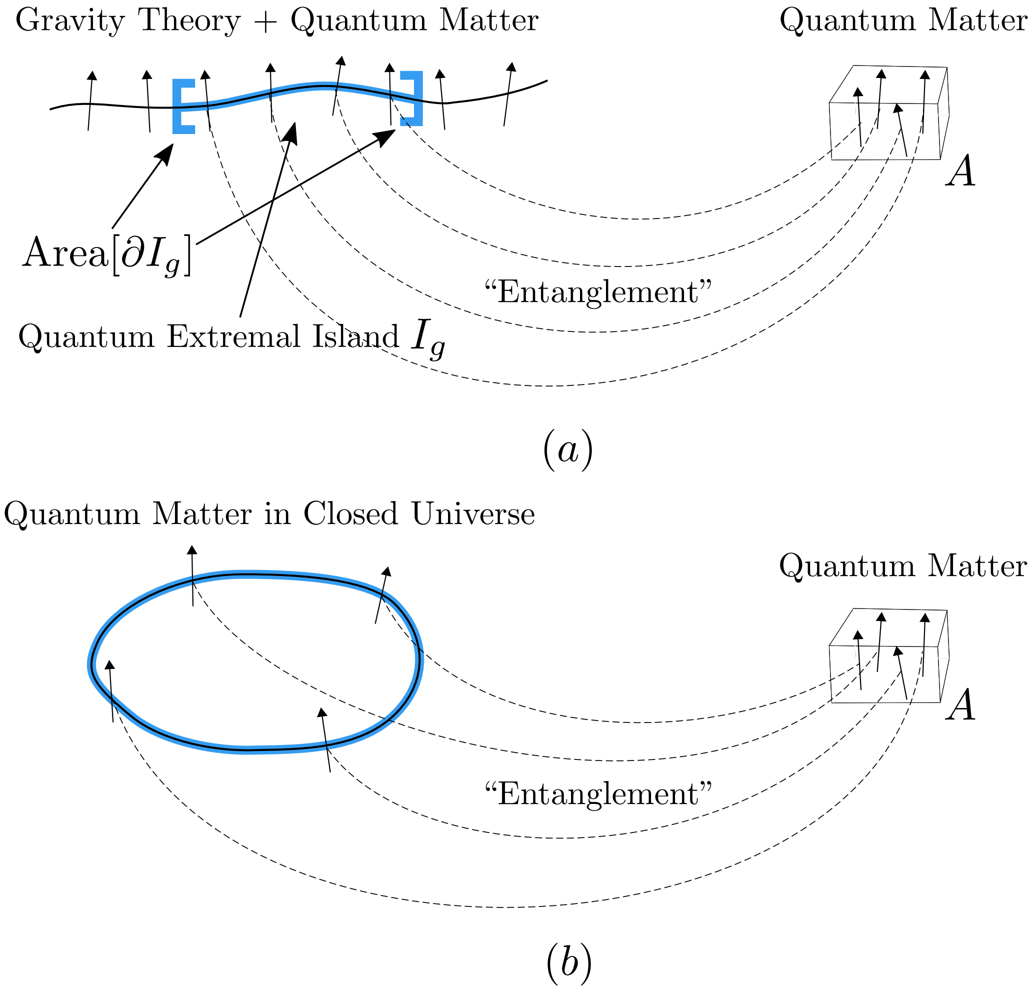


Image credit: [42]

Figure 3.4.1: A qualitative depiction of the quantum extremal islands proposal for the case of (1 + 1)-dimensional dilaton gravity coupled to a CFT_2 in the bulk, where the latter's contribution can be geometrically computed via its AdS_3 dual. This example was the one studied in the prescription's original formulation [42]. By thinking of it in a holographic context, one can separate the system (which as a whole renders a pure state) into a bulk gravitational theory coupled to the CFT (left) and a distant part of the boundary QFT region A (right). Dividing these two systems creates an entangled bipartite state with a generally non-vanishing von Neumann entropy (as seen by observers from either side). (a) shows the case of an island in an unbounded universe, while (b) shows that of a topologically closed one, where the extremal islands enclose it completely in order to extremize the entropy functional as per the maxi-min construction.

[53] Randall and Sundrum famously argued that the apparently four-dimensional nature of spacetime (with the Standard Model and the manifestly four-dimensional law of universal gravitation) could be the product of such a confinement, suggesting that the comparative weakness of gravity is a consequence of it being "leaked" to higher non-compact dimensions [55].

4 Recovering the Page Curve

We will now discuss how the methods from the previous section can be applied to obtain Page curves from radiating black holes.

4.1 A Prescription for Evaporating Black Holes

Strictly speaking, so far we have only explicitly dealt with computing the entropy of some holographic-like region where QFTs live. It is not *a priori* obvious that this should also comprise black holes. However, we have already seen a hint that it can: In particular, we can think of the picture in Fig. 3.4.1 in terms of a quantum-gravitational system at one end, coupled to a distant region whose quanta live in an effectively static spacetime. This should ring some bells, it is precisely the type of situation that Hawking originally considered: a black hole emitting radiation to a distant observer.

In this light, we can interpret the proposal (3.4.7) as the entropy an observer would measure by collecting Hawking quanta of a distant black hole and properly accounting for its gravitational origin. This latter point is a crucial one: One can't simply compute the entropy of an effectively random set of spins, this would indeed look thermal and lead to the violation of unitarity that Hawking postulated. The key insight of the islands proposal is that there will in general be disconnected regions (potentially far from the observer) that must be accounted. One can think of these as being connected by a higher dimension in the manner of ER = EPR [42, 56], i.e. the statement that entanglement is in some sense caused by a wormhole-like connection between particle pairs. It is this entropy that one would expected to obtain with full knowledge of the density matrix of the system and tracing over the non-radiation modes, i.e. it is the von Neumann fine-grained entropy of the system.

If we take the state to be pure, this also means that the entropy of such radiation should be equal to that of the (non-radiation) black hole system. This is what one would expect, although it is not good enough to just assume it away, as this is what we are to find evidence for (not the other way round). In particular, on simple grounds given the BH entropy and elementary knowledge of quantum fields, it is easy to argue that the entropy of a black hole should take the generic form

$$S_{\mathrm{bh}} = \frac{\text{Area of boundary}}{4G_N\hbar} + S_{\mathrm{ent}}(\text{exterior region}) + \text{corrections}, \tag{4.1.1}$$

which has an obvious similarity with Eq. (3.4.1). The significance of this similarity, on the other hand, is not trivial at all. It means that this whole machinery of prescriptions to calculate the entropy of QFTs at the boundaries of exotic spacetimes has convergently evolved to share much of the same qualities of what could be regarded as common knowledge of black holes. One could be suspicious that there has been more hindsight than foresight to these proposals, after all, the initial RT prescription was clearly based on the Bekenstein bound, so it is clear that from the start there was a motivation to connect the two. However, while there is some truth to this on a conceptual / pedagogical level, we are still left with a remarkable fact: the proposals work. We saw this explicitly when looking at how this approach yielded the correct boundary entropy in $\mathrm{AdS}_3/\mathrm{CFT}_2$, and there are many more instances of this. Note also that there is not a lot of wiggle room either when it comes to having a rigorous path integral prescription that correctly yields these proposals. So it seems that RT's original intuition was more than some "prejudice" on how Nature ought to behave, it was the anticipation of a deep connection between quantum and gravitational systems.

Thus, in the spirit of [42], we shall interpret the islands prescription (3.4.7) as the guiding principle for computing entropy in black hole systems, and for that matter, the Page curves associated to them.

The basic idea that the thermodynamics of black holes, as seen from the outside, should be the result of unitary evolution has a name, it is known as the "central dogma" and can be stated as follows [2]:

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

The term was coined in analogy with the central dogma of biology, which is a fundamental statement about the conservation of information in the DNA $\to$ RNA $\to$ protein transcription process, which is analogous to black holes in its significance and its relation to information.

By "degrees of freedom," this refers primarily to the logarithm of the dimensions of the Hilbert space, which are manifestly finite in this description. It should be interpreted flexibly in that it is not manifest in the gravitational description, although one could roughly think of it as the area term in (4.1.1). The area is not necessarily that of the horizon as in the BH entropy, as we will in general have to deal with islands that extremize the entropy functional. Also, although one would be tempted to say that the degrees of freedom of ingoing Hawking quanta get "transferred" to those of the black hole, this is a problematic statement, as the computation in the islands prescription includes disconnected regions inside the horizon for computing the radiation, as seen by a distant observer. There is also some ambiguity in what can be called the "exterior." In practice, if the system is not confined to a reflective region and we are considering asymptotically flat / static spacetimes , then one effectively defines an effective cut-off or "atmosphere" where all the relevant dynamics take place and call this the inner region (which should evolve unitarily).

We also see that this is not a statement about the interior, so it doesn't make too many assumptions on what should or should not happen inside it. Of course, if matter and radiation get destroyed in the singularity, then this would certainly be a cause of concern, since if their information gets erased, then unitarity would be violated. Black hole complementarity suggests, as seen from the outside, objects never reach the singularity. Unitarity also implies the existence of a Hamiltonian that generates the evolution of the system. It turns out that if one attempts to define full Cauchy slices of the system (including the interior) where such a Hamiltonian can be defined, it leads to the existence of slices where two copies of the system exist, which would violate the no-cloning theorem for quantum systems. This famously led to the idea of black hole complementarity [4, 57, 58], which postulates that no observer ever sees both copies of the system.

Among other things, it says that infalling objects take an infinite amount of time to reach the horizon (and thus the singularity), favouring unitarity.[54]

For the most part, we won't worry too much about these subtleties. We are not trying to rigorously state the dogma, nor is it assumed in any of the work discussed, it is simply a rough conceptual notion of how one should think about unitarity in black hole evaporation. From our point of view, the utility of these prescriptions that we have discussed lies in the fact that they generate independent evidence for the central dogma when applied to computing Page curves. For that matter, we should think of it more as a proposition than a dogma.

It is indeed a fundamental principle in physics that systems evolve according to deterministic equations of motion. The only major exception to this appears to be the measurement problem. However, we know that, with "many worlds" type interpretations, one can see the act of measurement as a unitary process that only appears to violate unitarity from the point of view of an observer that has no access to what may be called the "wave function of the universe." So once more, in many worlds, the apparent uncertainty in measurement arises due to restricted accessibility to the whole quantum state (just as in black holes), which is the more conventional context in which the density matrix approach is used.

4.2 The Entropy of a Black Hole

We will now review how this procedure is applied to computing the entropy of black holes, based on the work found in [43, 44] and later discussed in [2, 42].

Fig. 4.2.1 shows the procedure at work. The way it evolves for a typical evaporating black hole is as follows: Consider a black hole that forms via some unitary process (e.g. quantum condensation of a cloud of matter, bosons etc.). This is reflected by the extremal surface at the initial stages, where there is no radiation, meaning that the minimal entropy arises due to the trivial / vanishing surface, i.e. shrinking the area of the region to zero and computing the non-existent entanglement entropy outside it, thus rendering a pure state.[55] This is in contrast to the BH entropy, which is initially zero and then $4\pi r_s^2/G_N$ after the black hole forms, further corroborating Bekenstein's idea that it should be treated more as a bound than an intrinsic (non-unitary) property of matter.

As radiation starts being produced, ingoing modes build up inside the black hole horizon and get entangled with their partner ones beyond, increasing the entropy of the vanishing surface. At some non-trivial point, it will pay in the maxi-min prescription to expand the area, by allowing for a trade-off in the purification of entangled interior-exterior pairs. This is when we start to have a non-vanishing entropy, so we see that this arises due to the creation of these radiation modes.

Naively, one may think that the vanishing surface should always be the one that minimizes the entropy, after all, shouldn't a full Cauchy slice (up to the cut-off) purify

[54] There are two ways of seeing this: Nature forbids observers from seeing unitarity violated, so it isn't violated, or unitarity is indeed violated, but with no observable consequences. As far as science is concerned, this is not a meaningful difference. Philosophers may differ.

[55] We are ignoring the entropy that arises for adding a cut-off surface, this is taken to have a constant time-independent contribution, so it will not be relevant in our discussion.

all necessary quanta with zero cost to the area? This is not true for two reasons: one is that while one imagines radiation quanta pair production as a simultaneous event, this is not true for all reference frames, meaning that there exist Cauchy slices where ingoing quanta simply can't be purified in this way. The other reason is that the quanta beyond the cut-off, which are part of the radiation system, aren't included in the computation. Otherwise we would get a trivially pure state, as a Cauchy slice covering the entire space occupied by the black hole + radiation system, should indeed evolve unitarily. So to sum this up, it is sometimes best to pay some area cost to exclude those unpurifiable modes from the slice in the entanglement entropy computation.

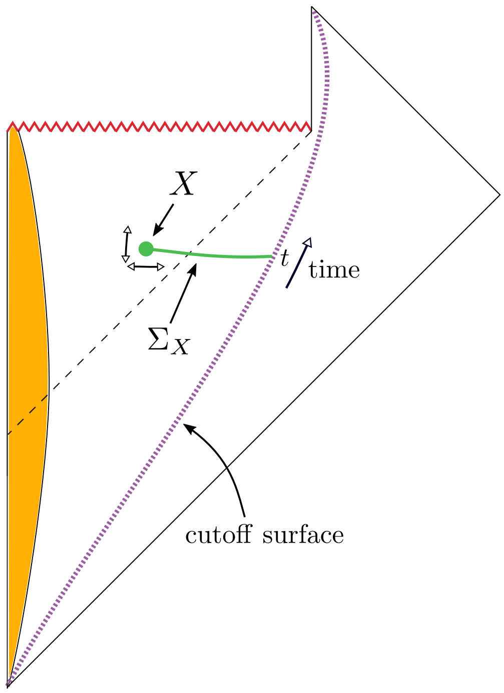


Image credit: [2]

Figure 4.2.1: Penrose diagram of an evaporating black hole formed by collapse showing how the generalized entropy prescription is applied to the case of finding a single quantum extremal surface. As usual, in $3+1$ dimensions, each point represents a two-sphere. One varies the position of the surface X that gives rise to the area term and computes the entanglement entropy of the exterior region Σ_X up to the cut-off of the Cauchy slice in order to extremize the black hole entropy functional.

Fig. 4.2.2 depicts how this build-up of unpurifiable quanta arises in the vanishing surface, making it manifestly unsustainable for a unitary process at later times. This means that while at earlier times, the dominant contribution to the extremal surfaces are the vanishing ones, at later ones there is a non-trivial surface that corresponds to this trade-off between the area cost and purifying entangled modes in the radiation. The nature of this latter phase involving the trade-off is explored in Fig. 4.2.3.

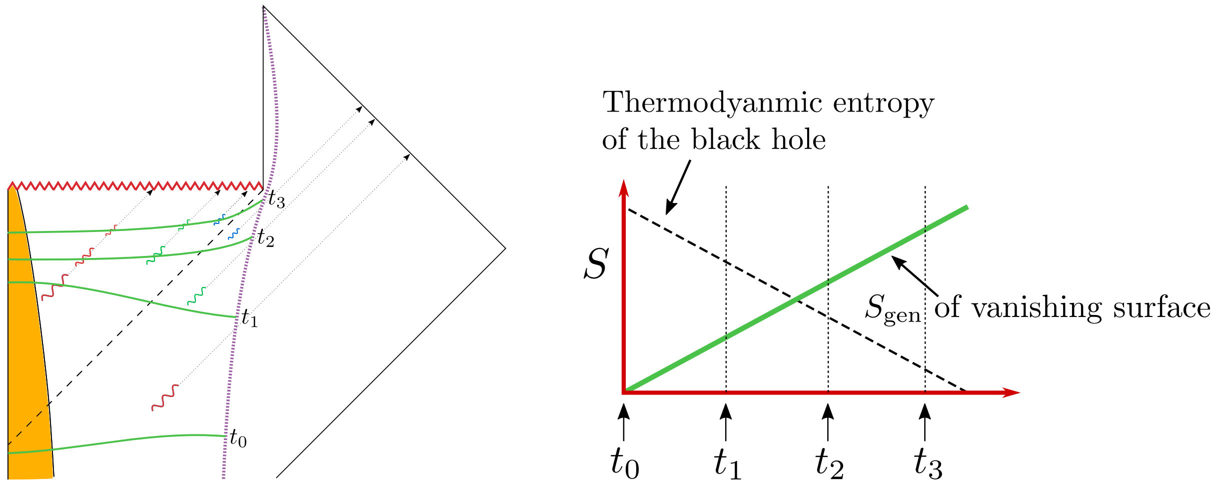


Image credit: [2]

Figure 4.2.2: In the left, we see the Cauchy slices of vanishing surfaces in the Penrose diagram of a black hole. The null trajectories represent those of pairwise-entangled radiation modes. We see how pairs don't necessarily lie in the same Cauchy slice, as they can be asynchronous or lie outside the cut-off boundary. At the time the event horizon forms t_0, no such radiation modes exist and the vanishing surface extremizes the entropy, corresponding to the initial condition of a momentaneously pure state everywhere in the system (not just as a whole, which is always true). As time progresses ($t_{1,2,3}$), the entropy of the vanishing surface steadily increases due to the accumulation of unpurifiable radiation, to the point it violates the Bekenstein (BH entropy) bound at $t_{2,3}$, implying the vanishing surface at later times can't be suitably extremal for a unitary process (right).

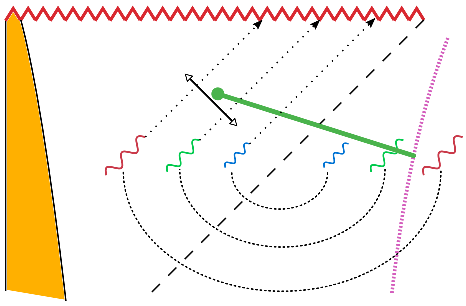


Image credit: [2]

Figure 4.2.3: The precise form of the extremal surface in the non-trivial scenario involving pairwise-entangled radiation can be summarized by the following simple algorithm: Start with a Cauchy slice with the boundary of the area term intersecting the horizon and extend it along the ingoing null direction until no further radiation can be purified. In this way, one arrives at a point where one can purify as many modes as possible without the superfluous unpurifiable ones that arise in the vanishing surface.

The trade-off described in Fig. 4.2.3 forms the key conceptual basis to prove the existence of a non-trivial extremal surface in this later regime when the vanishing surface doesn't yield a viable maxi-minimization. In particular, it is by exploiting the fact that, by moving in the ingoing direction of the horizon, the area term is monotonically decreasing, that this must at some point balance with the opposite contributions of including unpurifiable states. Along this lines, one can show that there exists a point about which any deformation of the surface exactly cancels, which precisely describes the behaviour of an extremal surface. This was done on more formal grounds in [43, 44].

It turns out that this extremal surfaces lies close to the horizon. Specifically, for a Cauchy slice at a time t, it will be found in the vicinity of the null trajectory connecting the horizon with the cut-off surface at an earlier time $\sim t - r_s \ln S_{\mathrm{BH}}$, where $r_s \ln S_{\mathrm{BH}}$ is known as the scrambling time[56] and is small compared to the evaporation time $r_s S_{\mathrm{BH}}$. This situation is explored in Fig. 4.2.4.

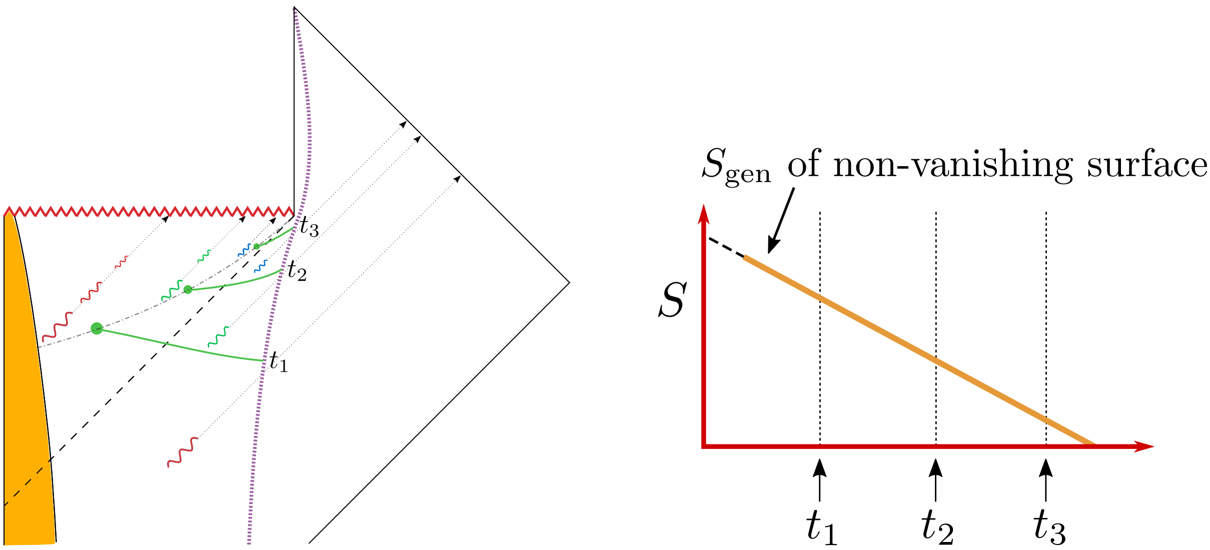


Image credit: [2]

Figure 4.2.4: The left shows the location of the non-vanishing surface at different times, which approaches the event horizon in an asymptotic manner as time elapses. The right shows how it steadily decreases with time as the event horizon shrinks due to evaporation, reducing the size of the surface. The limiting case occurs when the entire black hole evaporates, so that the surface (and thus the entropy) vanishes, resulting in a pure state entirely composed of Hawking radiation. This is much like in the beginning.

[56] This concept was introduced from a quantum information point of view in a (by now well-known) paper by Hayden and Preskill [59]. One can think of it as follows: Consider a system of entangled qubits. If one sends a portion of them into a black hole and then analyse the emitted radiation, one should be able to infer and effectively measure the state of the qubits that were thrown. The scrambling time is a measure of the time lag that arises due to the black hole absorbing those qubits, re-shuffling them in its interior, and emitting them back as part of an approximately uniform "scrambled" radiation. This behaviour is analogous to that of quantum mirrors, which was the original motivation behind the idea.

Due to the proximity of the non-vanishing surface to the horizon, we see that the area term dominates at this stage. This makes sense, as we would expect that at these late times, the comparatively high abundance of ingoing radiation modes means that one should not go too far inside the horizon to meet the trade-off point between the reduction in surface cost and penalization due to a a high number of unpurifiable states, whose pairs will have already crossed the cut-off surface by that time.

Since this dominating area term is close to the horizon, it also means that the entropy at this point, is about equal to the thermodynamic BH entropy, so that at late times:

$$S_{\text{non-vanishing}}(t) \approx \frac{A_{\text{ext}}(t)}{4G_N\hbar} \approx \frac{\text{Area of horizon}(t)}{4G_N\hbar} = S_{\text{BH}}(t), \tag{4.2.1}$$

where A_{ext} denotes the extremal surface and we have emphasized the time dependence due to the shrinking horizon at this stage due to evaporation.

It turns out that this transition between the vanishing and non-vanishing extremal surfaces is quite abrupt, meaning that as soon as the vanishing phase is no longer feasible, the non-vanishing one quickly takes over and approaches the horizon. This occurs roughly at the midpoint between the two phases, when the black hole becomes to shrink and the vanishing surface entropy saturates the Bekenstein bound. This should be clear, as we know that the later non-vanishing stage also saturates it and that the vanishing-surface, being monotonically increasing, would violate the Bekenstein bound (and thus unitarity) if extended further in time, so this is the necessary meeting point between the two regimes. Of course, what we are describing is the Page curve of the system, as shown in Fig. 4.2.5.

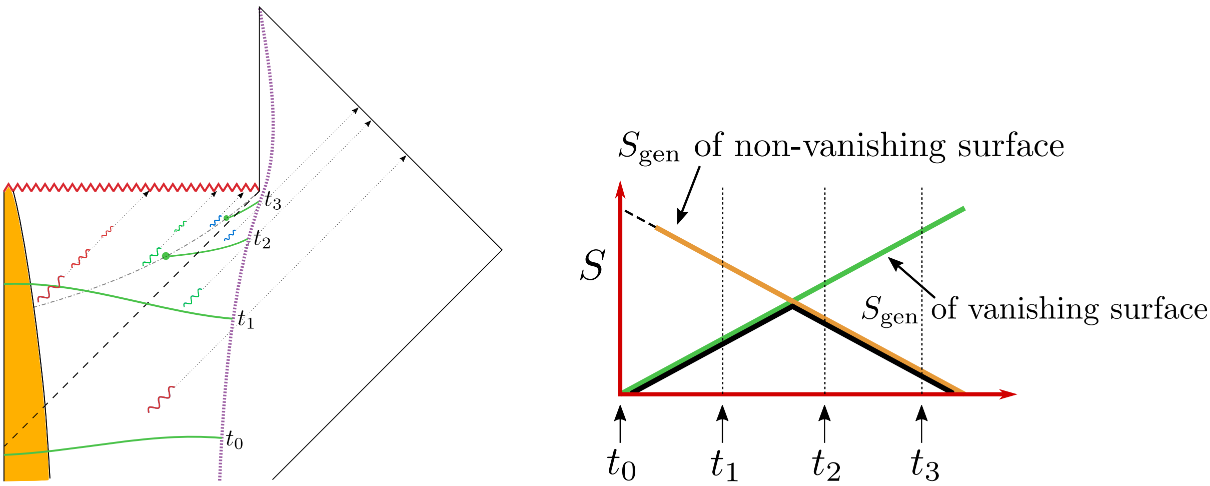


Image credit: [2]

Figure 4.2.5: Page curve of the system obtained from the piecewise combination of the vanishing and non-vanishing extremal surface regimes shown in Figs. 4.2.2 and 4.2.4 respectively. When the horizon forms at t_0, we have the vanishing surface rendering the pure state, which also remains dominant at t_1. At $t_{3,4}$, the system has made the swift transition to the non-vanishing surface phase, where the thermodynamic entropy bound is saturated thereafter, ultimately returning to a pure state at the evaporation time.

4.3 The Entropy of Radiation

While the above discussion on the evolution of the entropy of a radiating black hole is indeed consistent with unitarity, this does not directly address the Page curve and the entropy that an observer would measure, as this would be the entropy of the radiation. While for unitary evolution these two should be equal, this can't simply be assumed, as this would amount to a somewhat circular reasoning and leave the more relevant question of the entropy of the radiation uncertain.

There is indeed a valid concern that we should consider the increasing amount of entropy leaving the cut-off surface, which is what distant observers would measure in an approximately static spacetime. The cut-off surface can be seen in a similar manner to a partition dividing the whole system, with the caveat that disconnected regions are allowed in both sides, so one can't simply say that one side of the partition is the black hole and the other the radiation. For the case of the black hole, these disconnected regions didn't become extremal, so this would have not been a relevant consideration. This is where the island approach becomes especially relevant as, in this case, there will exist disconnected regions that are extremal to the entropy functional (3.4.7). This new situation is illustrated in Fig. 4.3.1.

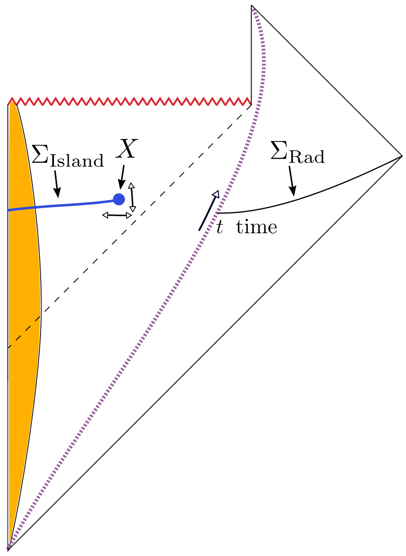
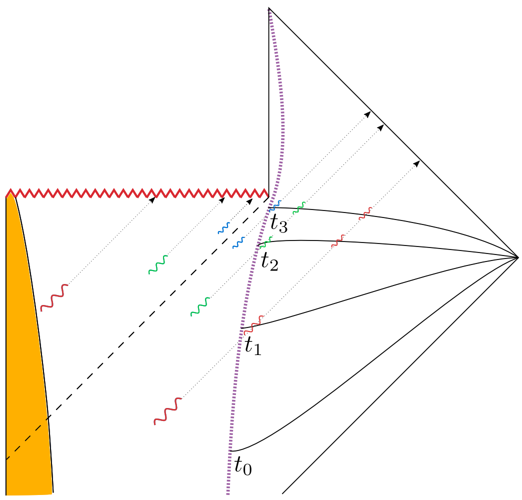


Image credit: [2]

Figure 4.3.1: Setup for computing the entropy of radiation with an island inside the black hole (cf. Fig. 4.2.1 in the black hole entropy case). The island involves an area term and thus has an associated cost to it. Islands arise when the cost of creating an extra area term is overcome by the purification of inner radiation modes in the black hole with the exterior ones in the main region outside the cut-off. Analogously to the black hole entropy, the initial regime comprises a vanishing surface in the complement exterior region that monotonically increases, until the area vs purification trade-off occurs, forming islands.

The fact that one isn't able to keep both systems neatly separated by the partition shouldn't be too surprising if one considers the following: that the black hole absorbs a radiation quanta for each pair that is produced near the horizon, meaning that there is a non-trivial distinction between what may be called the "black hole degrees of freedom" that are inherent to the system and the subsequent radiation modes that build up inside it. This illustrates the non-manifest behaviour of degrees of freedom in the gravitational description, unlike in the central dogma, which implicitly takes the somewhat oversimplified view that these should be easily distinguishable.

Although we could discuss at length the details involving the dynamics of the entropy from the radiation, to understand this, one only really needs to know the following: The extremal surface for the radiation at all times is simply given by the complement region to that of the black hole, thus making the two entropies equal by forming a globally pure state. At the initial stages, this takes the form of a surface covering the entire space beyond the cut-off region, which is indeed the complement of the vanishing surface. This is shown in Fig. 4.3.2. We can then do a piecewise combination of the two regimes in Figs. 4.3.2 and 4.3.3 to obtain the Page curve of the radiation as shown in Fig. 4.3.4.



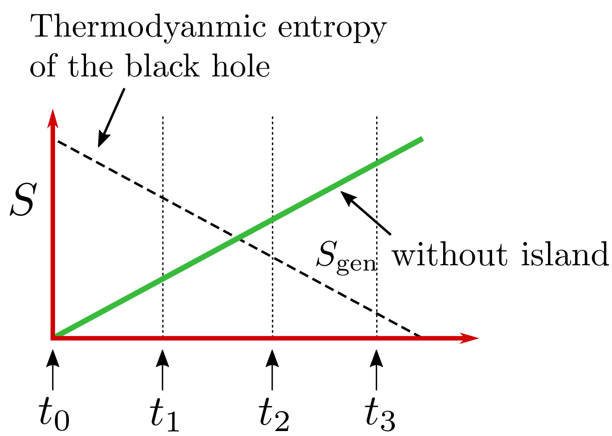


Image credit: [2]

Figure 4.3.2: The non-island contribution to the extremal surface. It is the complement region to the vanishing surface of the black hole shown in Fig. 4.2.2 and dominates at earlier times when it starts as a pure state, following the same behaviour. It is also vanishing in that it has no area term either. This region may be regarded as an approximately static spacetime with no black holes, so that the computation reduces to the amount of entanglement entropy of the radiation in this region (thus the lack of an area term).

At later stages on the other hand, we have the island contribution that is complement to non-vanishing surface in the black hole, this is shown in Fig. 4.3.3. We see how the two entropies behave in precisely the same way, saturating the Bekenstein bound. This reproduces the same result as for the black hole in Fig. 4.2.5, indeed showing that we have a unitarily evolving system where the Hawking radiation is coupled to the hole.

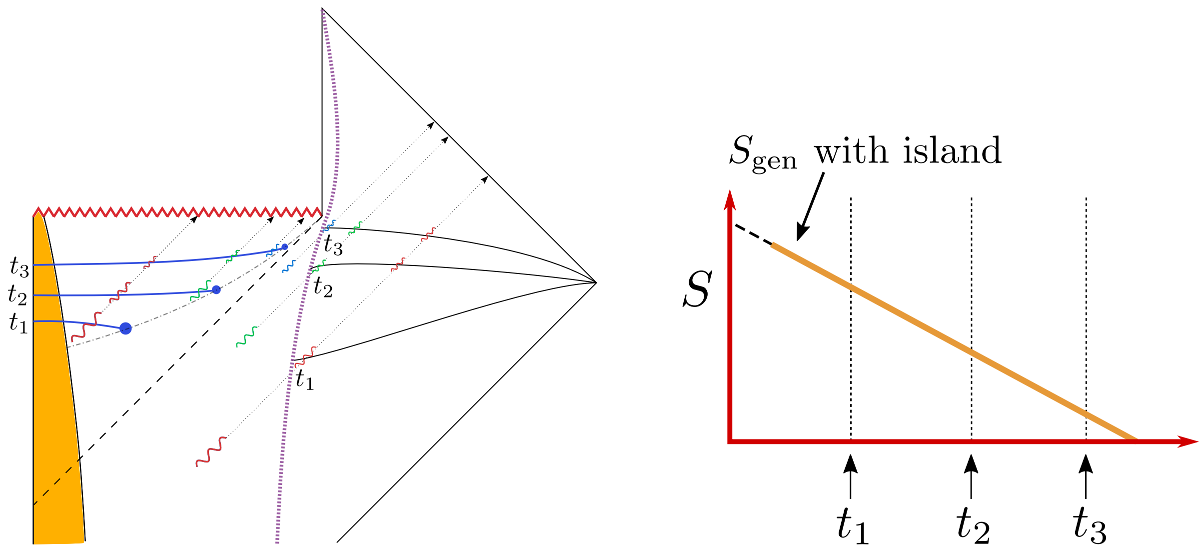


Image credit: [2]

Figure 4.3.3: The island contribution to the entropy. This regime is dominant at later times and consists of the two regions that are complementary to the non-vanishing surface of the black hole entropy in Fig. 4.2.4. This means that the area term in such surfaces is dominant and about equal to the BH entropy, while the remaining contribution is near-vanishing due to the purification of entangled Hawking quanta in the radiation.

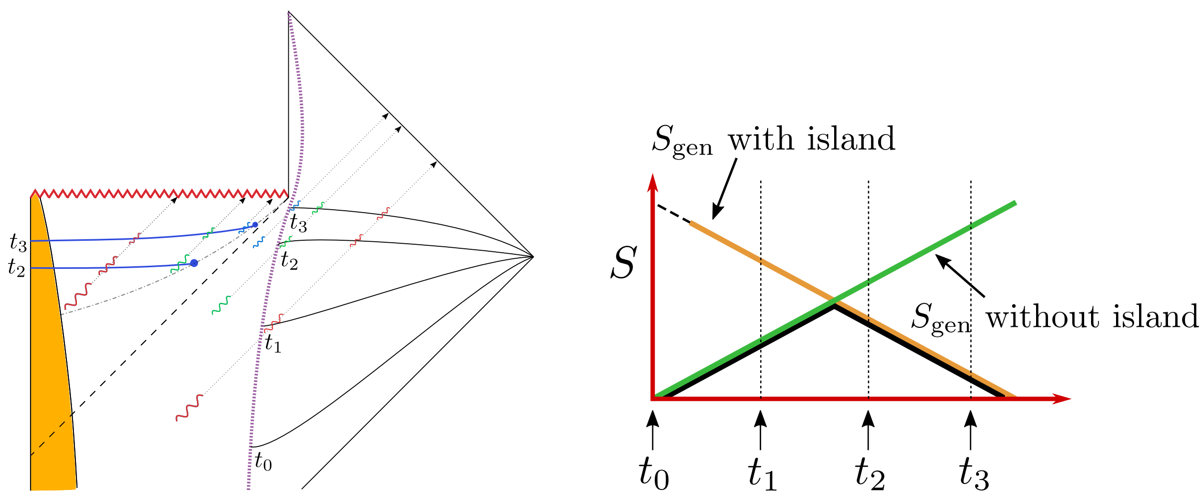


Image credit: [2]

Figure 4.3.4: The resulting piecewise combination of the two dominant regimes yielding the Page curve of the radiation. This precisely matches with the entropy of the black hole result in Fig. 4.2.5, in agreement with what is to be expected from unitarity. In particular, we see how the non-island vs island regimes exactly correspond to those of the vanishing vs non-vanishing surfaces in the black hole case (respectively).

4.4 The Entanglement Wedge

We see from the extremal surfaces approach that the entropy has a manifest dependence on the geometry of the black hole interior. There is also the more general question of what degrees of freedom does the entropy describe and how these depend on regions of the system at later times.

While this may sound like a potentially complicated question, the answer is quite simple: we construct the pass and future light cones that causally connect the extremal surface with everywhere else in the spacetime. This is known as the "causal diamond", as it is the shape it has in a spacetime diagram for a flat Cauchy slice, which is illustrated in Fig. 4.4.1.

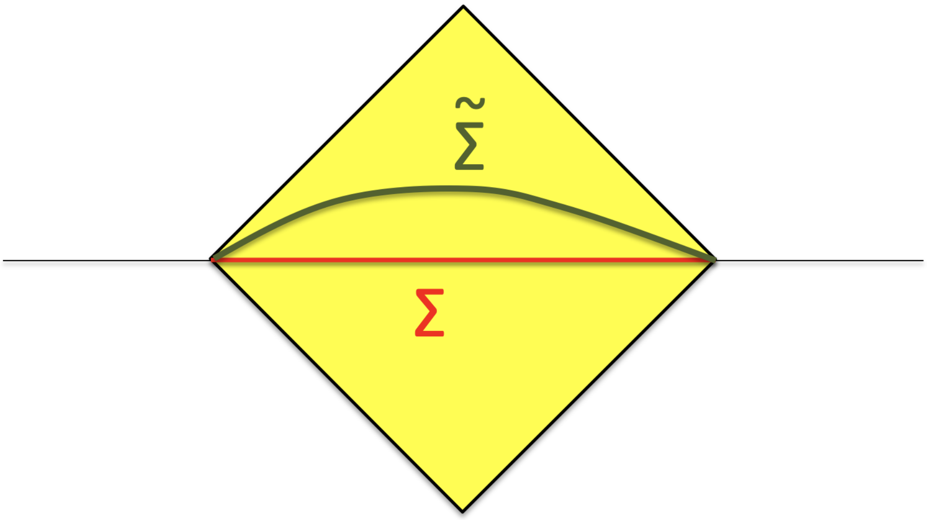


Image credit: [2]

Figure 4.4.1: Causal diamond of a flat surface Σ constructed by joining two light cones that define its causal influence on the spacetime. This means that any arbitrary surface $\widetilde{\Sigma}$ within the diamond has initial conditions fully determined by Σ. The number of such conditions has direct correspondence with and is bounded by the degrees of freedom (and thus the entropy) of Σ defining the region. This is highly analogous to the light-sheet construction in the Bousso bound (applied to the HRT proposal) that was discussed earlier in Section 3.3 (cf. Fig. 3.3.1).

When one applies this concept to the extremal surfaces of the fine-grained entropy functional (3.4.7), it is known as the "entanglement wedge" [2, 54, 60, 61].[57] This is shown in Fig. 4.4.2. The entanglement wedge defines all the regions of spacetime that are causally defined by the boundary conditions of the extremal surface. Such boundary conditions give a precise meaning to what we mean by "degrees of freedom." This detail

[57]This is a rather vague, but nevertheless predominant, terminology. A more accurate term that has been proposed is the "fine-grained entropy region" [2].

is more or less ignored in the central dogma, where one only considers the apparent degrees of freedom of the entire region inside the horizon as seen from the outside. In particular, one should not be misled by the fact that "black hole degrees of freedom" in the semi-classical gravity description (3.4.7) does not include the inner radiation modes inside the horizon, while the central dogma does not make this distinction.

This lays the resolution to a long-standing problem proposed by Wheeler [2, 62] concerning the seemingly paradoxical situation where one could have a bottle-neck-like geometry enclosing an arbitrarily large amount of entropy that far exceeds the Bekenstein bound. Such geometries are known as "bags of gold." The solution to this is simple: The highly entropic region does not lie in the entanglement wedge, and so the bound does not apply, as it only describes the entropy in the vicinity of a comparatively small horizon.

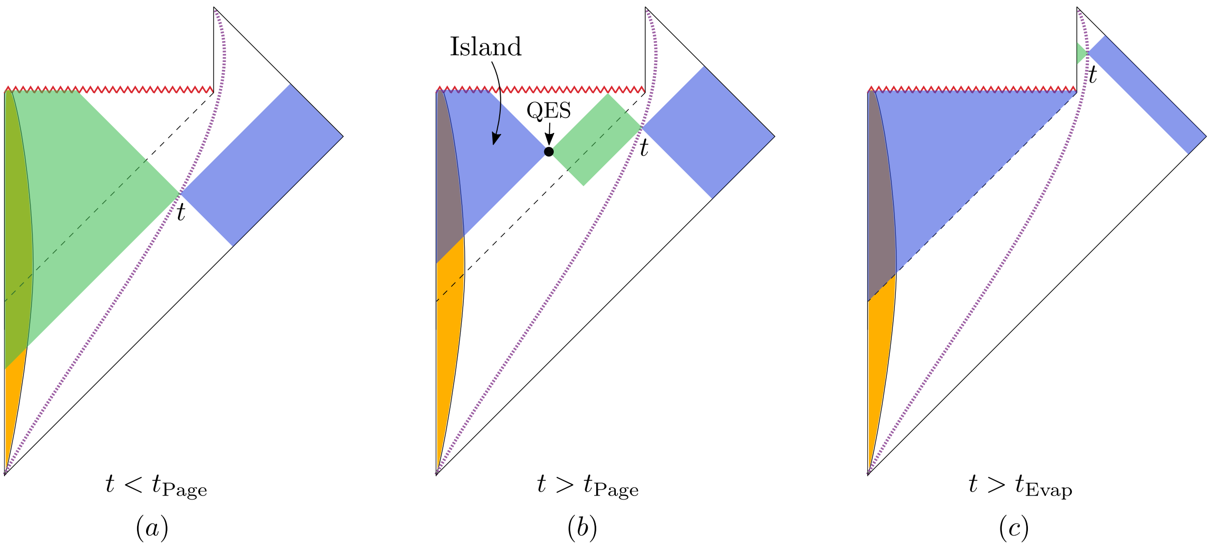


Image credit: [2]

Figure 4.4.2: Entanglement wedges for black hole evaporation of the black hole (green) and the radiation (blue) systems. This is characterized by three stages:

(a) Initially, before the Page time, we have the vanishing surface for the black hole whose wedge covers a large portion of its interior and some of the spacetime beyond the horizon and the cut-off. The radiation covers only a portion of the spacetime outside the cut-off.

(b) After the Page time (but before the evaporation time) islands form and the radiation wedge now covers some of the black hole interior too, while the now non-vanishing extremal surface of the black hole entropy gets confined to a smaller region near the horizon.

(c) When the black hole evaporates, it ceases to exist thereafter. The entire spacetime of the black hole interior is covered by the radiation entanglement wedge. All that remains from the black hole are Hawking quanta encoding the totality of its information, from its formation to the entire history of its evolution.

5 Two-dimensional JT Gravity

By now, we have discussed all the key ingredients to understand the evolution of the entropy in black hole evaporation. It is now time to review how it all comes together in the case that originally formed the basis for this understanding, two-dimensional dilaton gravity coupled with holographic matter [42, 43, 44]. In particular, we shall consider a theory with the $(1+1)$-dimensional dilaton gravity action (3.4.2) coupled with a CFT_2 known as Jackiw-Teitelboim (JT) gravity, so that

$$I[g_{ij}^{(2)}, \phi] = I_{\text{grav}}[g_{ij}^{(2)}, \phi] + I_{\text{CFT}}[g_{ij}^{(2)}, \chi], \tag{5.0.1}$$

where χ collectively denotes the fields in the CFT_2. In essence, this is gravity coupled to matter with a higher-dimensional gravity dual, the dual just being the familiar case of gravity in the AdS_3 bulk for a CFT_2. To allow for to semi-classical treatment, one needs the CFT central charge to satisfy $1 \ll c \ll \phi/(4G_N^{(2)})$, which also ensures that the gravity dual has a large radius of curvature, making it more amenable to be studied.[58] In addition to this, the strongly coupled CFT limit is taken too, ensuring that the gravity dual will correspond to classical Einstein gravity.

The fact that we are dealing with holographic theories means that there will be an equivalence class of systems that are related by their duals, as shown in Fig. 5.0.1.

The so-called Planck brane in Fig. 5.0.1 arises from the analogous behaviour of this system to the Randall-Sundrum model [55] that we mentioned previously. In this two-dimensional form, one can find the precise embedding of the boundary by considering the behaviour of the stress-energy tensor, which can be obtained as a solution to (5.0.1),[59] and a two-dimensional metric living in the embedding space. The location along the one-dimensional line is often parametrized by $\sigma_y = (y^+ - y^-)/2$, where $y^\pm$ are suitably defined coordinates that are used to fully specify the Planck brane, so that any point along its surface has coordinates of the form (y^+, y^-) [42]. The parameter σ_y is a measure of the distance along the boundary line of the brane so that $\sigma_y < 0$ in the dynamic part of the boundary and $\sigma_y > 0$ in the CFT bath, meaning that the thick dot in Fig. 5.0.1 is effectively the origin of this coordinate. This is analogous to the cut-off surface discussed in the previous section.

In this picture, we can consider a black hole living in the coupled gravity region $(\sigma_y < 0)$ of the theory that slowly evaporates into the CFT bath $(\sigma_y > 0)$ that is akin to the static spacetime where a distant observer collecting Hawking radiation would live in.

In the spirit of the FLM/EW prescription (3.4.1), one obtains the generalized entropy functional to be extremized [42]

$$S_{\text{gen}}(y) = \frac{\phi(y)}{4G_N^{(2)}} + S_{\text{Bulk-}2d}[\mathcal{I}_y], \tag{5.0.2}$$

where y is a coordinate parametrizing the location along the two-dimensional bulk and $\mathcal{I}_y$ is an interval from that point to the point where the effective gravitational constant

[58] This can be seen from Eq. (3.2.4), where we saw how the central charge in CFT_2 is proportional to the AdS_3 radius of the bulk gravity theory which is dual to.

[59] The standard process comprises the application of Noether's theorem, from which one finds the conserved charges that define the stress-energy tensor based on the symmetries of the system.

vanishes at $\sigma_y = 0$. We have also used $\hbar = 1$ units and will continue to do this for the remainder of this section. Note that the first term is indeed the area term: the area of a point is some constant that can be absorbed into ϕ and can be interpreted an $A/(4G)$ term with the effective gravitational constant $G \equiv G_N^{(2)}/\phi$.

In this context, the dilaton field ϕ is also known as the "entropion field" [42]. Intuitively, the higher the values of this field, the lower the effective gravitational coupling and so the greater the degrees of freedom of the system, as its dynamics become less constrained.[60]

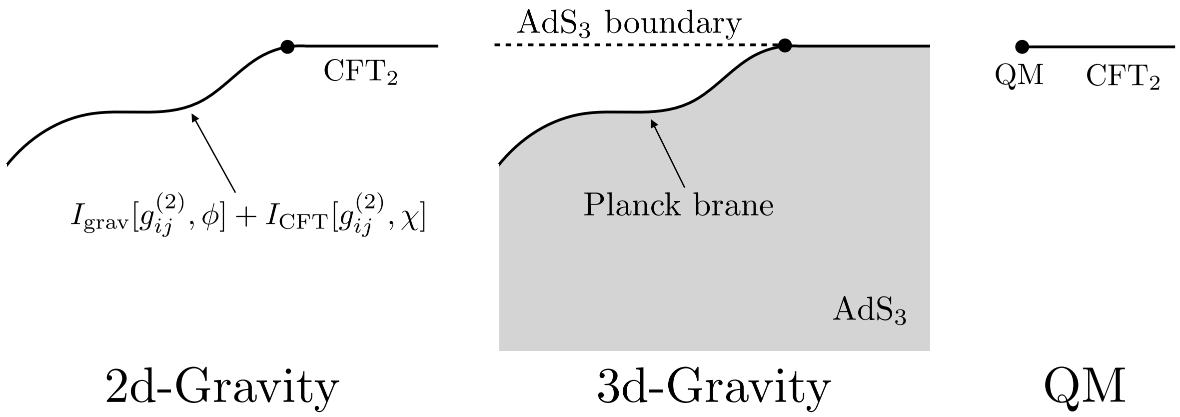


Image credit: [42]

Figure 5.0.1: Three equivalent versions of the gravity + holographic matter system. At the left we have a one-dimensional space region of two-dimensional dilaton gravity coupled with a CFT$_2$. At the thick dot, ϕ diverges so that the effective gravitational constant $G_N^{(2)}/\phi$ vanishes, meaning that the region beyond is a static spacetime with the equivalent of a CFT$_2$ "reservoir" or "thermal bath." In the middle, we replace the CFT$_2$ by its AdS$_3$ gravity dual, i.e. we think of the system in the left as living in an AdS$_3$ boundary. This boundary has a dynamical part known as the "Planck brane" and a static one which acts as the usual conformal boundary. In the right, assuming that the dilaton gravity theory has a $(0+1)$-dimensional dual, the system reduces to a quantum matter (QM) system consisting of a dual quantum theory + CFT$_2$ reservoir system with no gravity. In this case, the dual quantum theory resembles a QFT living in a point.

We can now turn this into a purely geometric problem by taking the AdS$_3$ dual of the CFT$_2$, which gives [42]

$$S_{\rm gen}(y) \approx \frac{\phi(y)}{4G_N^{(2)}} + \frac{\text{Area}^{(3)}[\Sigma_y]}{4G_N^{(3)}}, \tag{5.0.3}$$

[60] This is much like say, the gravitational constant with a gas of particles in a box. Normally, one has an approximately ideal gas of freely moving particles with a large number of degrees of freedom. However, if one could somehow increase the gravitational coupling sufficiently, these would clump together and form a highly constrained state in the phase space of the system, with little degrees of freedom.

where Σ_y is the usual surface in the AdS_3 bulk homologous to $\mathcal{I}_y$ and the approximation sign is needed to account for minor corrections in sub-leading terms, as well as possible fluctuations in the metric and the dilaton / entropion field. This means that, up to these small corrections, the problem has been reduced to one that is solvable via the more familiar RT/HRT prescription of extremizing areas, as seen in Fig. 5.0.2.

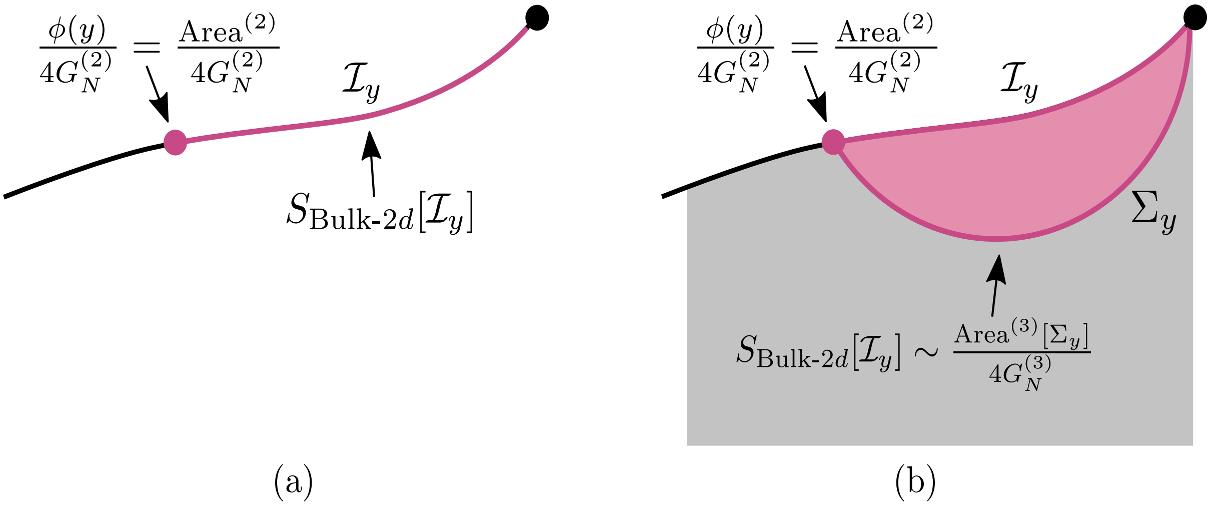


Image credit: [42]

Figure 5.0.2: Generalized entropy surfaces in (a) the interval $\mathcal{I}_y$ living in the $(1+1)$-dimensional gravity + CFT_2 theory, and (b) the version of this theory where we take the AdS_3 dual of the CFT_2. In this latter case, the entropy is obtained of the boundary $\mathcal{I}_y$ is obtained by extremizing its homologous surface Σ_y. This is essentially the same kind of problem as the one discussed in the RT framework earlier in Section 3.2 for AdS_3/CFT_2.

By establishing the above framework, one can now analyse the dynamics of the entropy in the system and derive its Page curve. We shall start by looking at the more instructive case of the later stage past the Page time. This situation is shown in Fig. 5.0.3. This is the analogue of the late stage non-vanishing extremal surface of the black hole entropy functional that we discussed in the earlier section. It too lies in the vicinity of the interior side of the horizon. In particular, the time-like coordinate from the endpoint y_e of the extremal surface is given by [42, 43, 44]

$$y_e^+ = t - \frac{1}{2\pi T(t)} \ln \frac{S_{\text{Bek}}[T(t)] - S_0}{c} + \dots, \tag{5.0.4}$$

where $S_{\text{Bek}}(T)$ is the Bekenstein entropy bound (i.e. the BH entropy of a horizon enclosing a region of spacetime, as discussed in Section 1.3) as a function of the temperature, which in turn has a time dependence due to the evaporation. S_0 is the extremal entropy, which we shall take to be comparatively small for the remainder of our discussion.

Eq. (5.0.4) was one of the key findings of [43, 44]. It precisely illustrates what was previously discussed on how the late-time non-vanishing surface is located near the point where an ingoing light ray would intersect the horizon if sent from the cut-off at a time earlier than t by a difference of order the scrambling time.

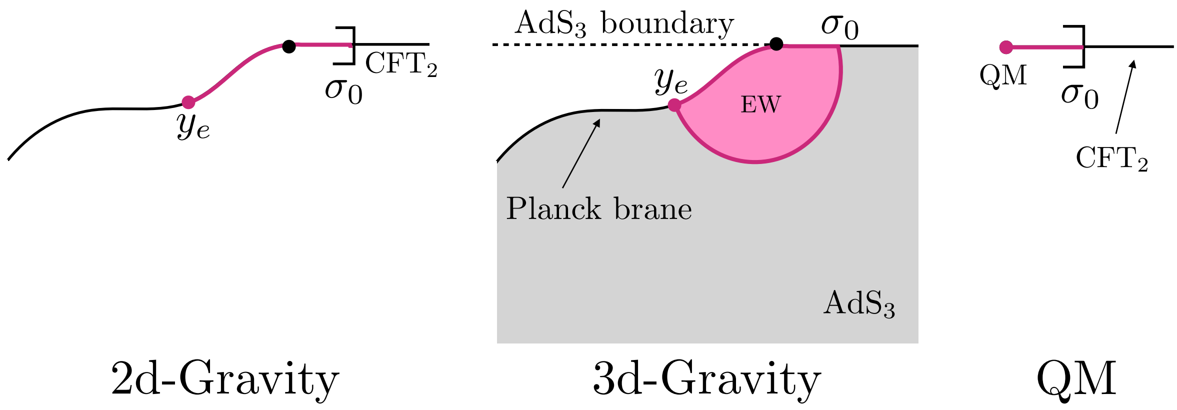


Image credit: [42]

Figure 5.0.3: Setup for computing the late-time entropy of a black hole system living in the theory, as interpreted in the three dual cases shown in Fig. 5.0.1. In the 2D and 3D gravity cases the region of the extremal surface is bound by an extremal endpoint just under the horizon $y_e = (y_e^+, y_e^-)$ and a another one (t, σ_0) in the CFT bath just beyond the cut-off point. t defines the time at which the entropy is computed. In the 3D case we have a section of the entanglement wedge of the black hole system as seen from a late-time Cauchy slice. In the QM case the extremal surface is in the interval $[0, \sigma_0]$.

It was then shown in [43, 44] that, the late-time black hole entropy is given by

$$S_{\text{Black hole}}(t) = S_{\text{Bek}}[T(t)] + \text{sub-leading logarithmic corrections}, \tag{5.0.5}$$

which indeed agrees with our discussion on how the late-time behaviour should saturate the Bekenstein entropy bound.

Now let's consider what is arguably the most interesting case examined in this analysis, the late-time behaviour of the radiation entropy. This is shown in Fig. 5.0.4.

We see how the islands approach really shines in Fig. 5.0.4, it establishes the existence of disconnected regions that must be accounted for, the islands. This leads to a non-trivial aspect of the calculation: One can't simply compute the entropy of a region while tracing everything else out, this is what would happen if we looked at the distant radiation only without accounting for the inner modes associate it it inside the event horizon. The result would be that we get Hawking's original result [6], which violates unitarity if taken on face value. On the other hand, the correct value by properly accounting for this matches that of Eq. (5.0.5), i.e. $S_{\text{Radiation}}(t) \sim S_{\text{Black hole}}(t)$, which indeed agrees with what is expected from the Page curve and unitarity.

We also see that while these regions appear disconnected and somewhat contrived to give the right results, perhaps nothing more than an "accounting trick," these are, on the other hand, manifestly connected via the AdS bulk in the 3D graviational description by taking the dual of the CFT$_2$. This key result traces us back to the remarkable utility of AdS/CFT to understand and formulate proposals of this kind. As we mentioned, this led to the interpretation that these two regions are an example of the ER = EPR conjecture. It was also proposed in [42] that this should extend to higher dimensions too.[61]

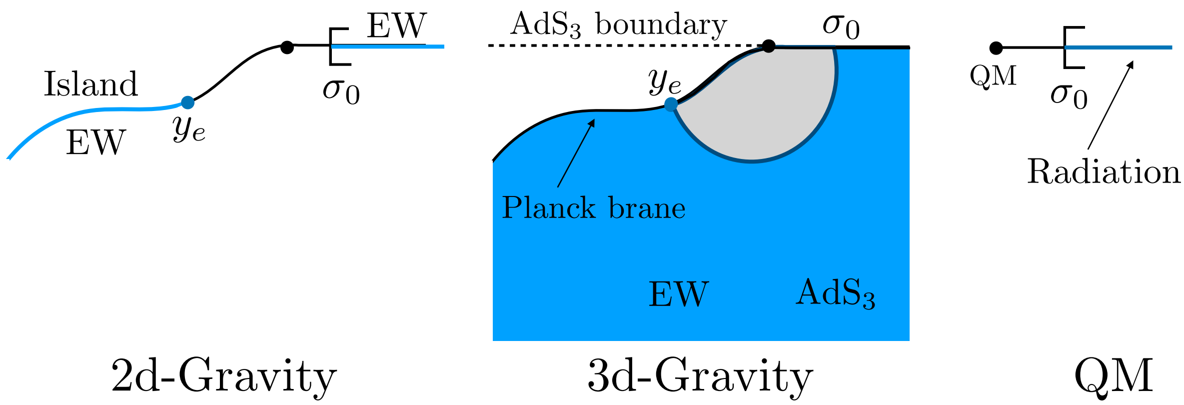


Image credit: [42]

Figure 5.0.4: Setup for computing the late-time entropy of the radiation system living in the theory, also as interpreted in the three dual cases shown in Fig. 5.0.1. This is simply given by the complement region to that of the black hole entropy in Fig. 5.0.3. Remarkably, what seems like two disconnected regions in 2d gravity (one in the interior and another one in the exterior bath) are in fact connected to each other via the AdS$_3$ bulk in the 3d gravity description. This was used to argue in [42] that it corresponds to a realisation of the ER = EPR conjecture [56], namely that distant entangled states that appear disconnected from each other are in fact connected through a higher dimension.

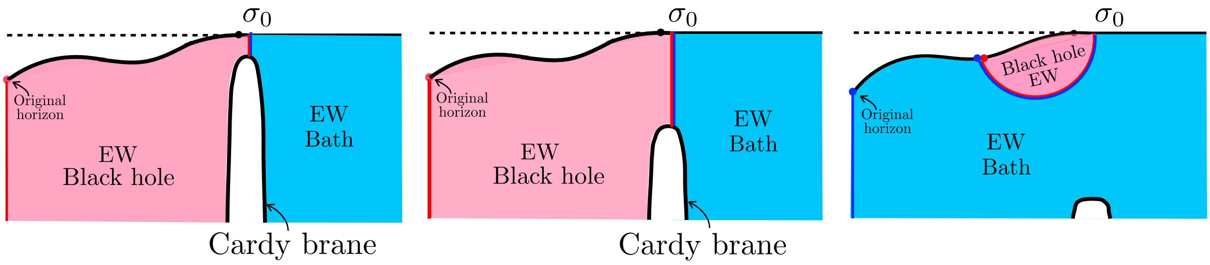
We will now consider the early-time behaviour of the entropy. This is a somewhat more complex in that it is more sensitive to the geometry and initial conditions. However, it is worth discussing to understand how this stage of the Page curve manifests itself in the system.

This is done by thinking of the CFT bath and the coupled gravity system with a black hole as two initially independent systems that do not interact with each other. Geometrically, this corresponds to creating a gap near the cut-off point which acts as some kind of impenetrable partition in a similar manner to dividing two gases in a box when considering them as thermodynamic systems. We can then "activate" the interaction at a time $t = 0$ and allow the two systems to interact from then onwards.

The problem with this is that this instantaneous change in boundary conditions is associated to a pulse of infinite energy, which is clearly unphysical. This means that one needs to allow for a finite time Δt to make the transition from non-interacting to interacting. The energy of the pulse in this case has an approximate energy $E \propto \frac{c}{\Delta t}$.

[61]Provided that evaporating black hole solutions exist, since in the Randall-Sundrum model [55], black holes generally have various phases that may or may not allow for evaporation.

The solution then lies in what the authors in [42] called “Cardy branes,” which consists in modifying the geometry of the usual bulk by creating an effective partition that allows the systems to interact gradually in a finite time, without the pathological behaviour inherent to infinite energy pulses. The name comes from the fact the theory obeys what are known as conformal Cardy boundary conditions [42, 63]. The resulting setup is explored in Fig. 5.0.5.



(a) $t = 0$ (b) $0 < t < t_{\text{Page}}$ (c) $t > t_{\text{Page}}$

Image credit: [42]

Figure 5.0.5: Dynamics of the JT gravity system and its extremal surfaces in the Cardy brane model assuming an initial low (non-zero) temperature black hole. The evolution of the system is characterized as follows:

(a) Initially, the black hole and bath systems have been restrained from interacting and their respective entanglement wedges extend everywhere up to the Cardy brane dividing them. The extremal surfaces enclose the entirety of the system in their respective sides of the partition. This situation is analogous to the vanishing surface characterizing the initial stages of the Page curve.

(b) For later stages earlier than the Page time, the distance between the Cardy brane and the cut-off boundary increases, while the entanglement wedges retain their basic topology. Physically, this increase in boundary length corresponds to the increase in entropy of the the vanishing surface due to the accumulation of trapped radiation quanta entangled with their exterior partners.

(c) At late times the Cardy brane no longer plays a significant role and we recover the behaviour discussed earlier in Figs. 5.0.3 and 5.0.4. We see the reservoir-like behaviour of the CFT bath, whose domain of influence, as characterized by the entanglement wedge, expands and plays a dominant role at late times. The limiting case of this is when the black hole completely evaporates and we are left with a “bath” of radiation with no black hole.

The early-time entropy corresponding to Fig. 5.0.5 (b) is given by [42]

$$S_{\text{Black hole}}(t) \sim S_{\text{Radiation}}(t) = \frac{\pi c}{6} \int_0^t dt' \, T(t') = 2S^i_{\text{Bek}}(1 - e^{-\frac{\kappa}{2}t}), \tag{5.0.6}$$

where (again) $T(t')$ is the time-dependent temperature. S^i_{Bek} is the initial Bekenstein bound of the black hole after absorbing the pulse generated due to the dynamics of the Cardy brane. The temperature of the excited black hole is taken to be much larger than that of the initial. It takes the form $T(t) \sim T_i e^{-\frac{\kappa}{2}t}$, where T_i is the initial temperature of the black hole after absorbing the pulse. κ is a constant proportional to c and the effective gravitational coupling in two dimensions.

Note that the expression for the entropy in (5.0.6) asymptotically tends to a value of $2S^i_{\text{Bek}}$ as $t \to \infty$. The factor of two is associated to the production of coarse-grained (thermodynamic) entropy in non-adiabatic processes. The corollary of this is that one should only take this early-time behaviour to be valid up to the Page time, where the original entropy bound S^i_{Bek} is saturated. It also means that the contribution of the leftmost RT/HRT surface extending vertically along the original horizon in Fig. 5.0.5 has a contribution to the entropy that can be ignored. This means that Fig. 5.0.5 (a) indeed corresponds to the vanishing surface with zero initial entropy. We can now summarize all the dynamics we have discussed by plotting the Page curve, which is shown in Fig. 5.0.6.

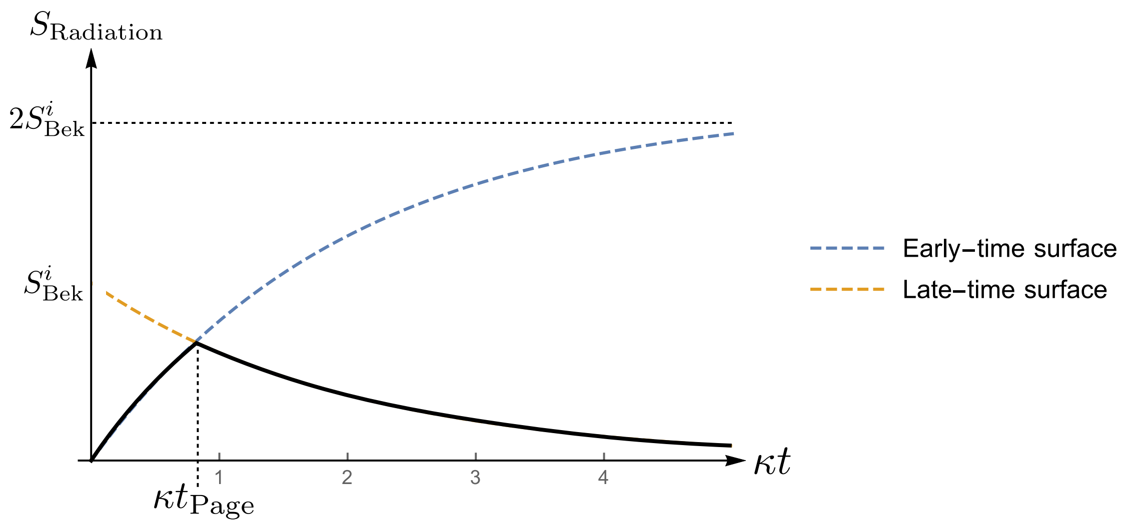


Image credit: [42]

Figure 5.0.6: Page curve of JT gravity (cf. Figs. 4.2.5 and 4.3.4). We see all the key stages reflected in the figure. We start with an initially pure black hole system that has no radiation and a vanishing extremal surface of zero entropy. The entropy of trapped entangled modes subsequently builds up, increasing by equal measure the entropy of the black hole and radiation systems, which form a collectively pure bipartite state at all times and thus have the same entropy. At the Page time, the vanishing surface saturates the Bekenstein bound of the initial state, this is defined to be the point where the entropy in the increasing early-time stage matches its decreasing late-time counterpart. Past the Page time, the entropy monotonically decreases due to the non-vanishing surface and an island formation in the black hole and radiation systems (respectively) due to the purification of entangled radiation pairs. In the final stage, only radiation remains after complete evaporation and the system forms a pure state made solely of Hawking quanta.

6 Further Comments

There are many possible applications and future directions to this work. One would be extending the various formulations of the entropy prescriptions or finding other supporting evidence for existing new ones like the islands proposal. Two-dimensional gravity models like the one discussed have been very instructive cases indeed, although one would like to see this extended in other more realistic models, perhaps so that it may derive direct conclusions on astrophysical spacetimes. Experimental progress with analogue black holes has also been reported [64, 65, 66].

7 Acknowledgements

I am very grateful to Fundación MAPFRE Guanarteme for their generous support in funding my studies at Imperial. I am also very thankful to Toby Wiseman for the very helpful discussions and guidance throughout the project. Lastly, I much appreciate Paul Campbell, my former teacher, for his much appreciated support and enthusiasm throughout the year.

References

[1] Stephen W Hawking. Particle creation by black holes. In *Euclidean quantum gravity*, pages 167–188. World Scientific, 1975.

[2] Ahmed Almheiri, Thomas Hartman, Juan Maldacena, Edgar Shaghoulian, and Amirhossein Tajdini. The entropy of hawking radiation. *arXiv preprint arXiv:2006.06872*, 2020.

[3] W. G. Unruh. Notes on black-hole evaporation. *Phys. Rev. D*, 14:870–892, Aug 1976.

[4] Daniel Harlow. Jerusalem lectures on black holes and quantum information. *Reviews of Modern Physics*, 88(1):015002, 2016.

[5] Richard C. Tolman. On the weight of heat and thermal equilibrium in general relativity. *Phys. Rev.*, 35:904–924, Apr 1930.

[6] Stephen W Hawking. Breakdown of predictability in gravitational collapse. *Physical Review D*, 14(10):2460, 1976.

[7] Don N Page. Information in black hole radiation. *Physical review letters*, 71(23):3743, 1993.

[8] Jacob D Bekenstein. Black holes and entropy. *Physical Review D*, 7(8):2333–2346, 1973.

[9] Claude Elwood Shannon. A mathematical theory of communication. *The Bell system technical journal*, 27(3):379–423, 1948.

[10] S. W. Hawking. Gravitational radiation from colliding black holes. *Phys. Rev. Lett.*, 26:1344–1346, May 1971.

[11] Jacob D. Bekenstein. Generalized second law of thermodynamics in black-hole physics. *Phys. Rev. D*, 9:3292–3300, Jun 1974.

[12] Jacob D Bekenstein. Universal upper bound on the entropy-to-energy ratio for bounded systems. *PHYSICAL REVIEW D*, 23(2), 1981.

[13] Leonard Susskind. The world as a hologram. *Journal of Mathematical Physics*, 36(11):6377–6396, 1995.

[14] Gerard'T Hooft. Dimensional reduction in quantum gravity. *arXiv preprint gr-qc/9310026*, 1993.

[15] Andrew Strominger and Cumrun Vafa. Microscopic origin of the bekenstein-hawking entropy. *Physics Letters B*, 379(1-4):99–104, 1996.

[16] Michael A Nielsen and Isaac L Chuang. *Quantum Computation and Quantum Information.* Cambridge University Press, 2010.

[17] Mark Srednicki. Entropy and area. *Physical Review Letters*, 71(5):666, 1993.

[18] Juan Maldacena. The large-n limit of superconformal field theories and supergravity. *International journal of theoretical physics*, 38(4):1113–1133, 1999.

[19] Juan Maldacena. The gauge/gravity duality. *arXiv preprint arXiv:1106.6073*, 2011.

[20] Sabrina Pasterski, Shu-Heng Shao, and Andrew Strominger. Flat space amplitudes and conformal symmetry of the celestial sphere. *Physical Review D*, 96(6):065026, 2017.

[21] Erin Crawley, Noah Miller, Sruthi A Narayanan, and Andrew Strominger. State-operator correspondence in celestial conformal field theory. *arXiv preprint arXiv:2105.00331*, 2021.

[22] Ana-Maria Raclariu. Lectures on celestial holography. *arXiv preprint arXiv:2107.02075*, 2021.

[23] Alexandre Belin, Jan De Boer, Jorrit Kruthoff, et al. Comments on a state-operator correspondence for the torus. *arXiv preprint arXiv:1802.00006*, 2018.

[24] Idse Heemskerk, Joao Penedones, Joseph Polchinski, and James Sully. Holography from conformal field theory. *Journal of High Energy Physics*, 2009(10):079, 2009.

[25] Marc Henneaux and Claudio Teitelboim. Asymptotically anti-de sitter spaces. *Communications in Mathematical Physics*, 98(3):391–424, 1985.

[26] Tom Banks, Michael R Douglas, Gary T Horowitz, and Emil Martinec. Ads dynamics from conformal field theory. *arXiv preprint hep-th/9808016*, 1998.

[27] CP Burgess and CA Lütken. Propagators and effective potentials in anti-de sitter space. *Physics Letters B*, 153(3):137–141, 1985.

[28] Shinsei Ryu and Tadashi Takayanagi. Holographic derivation of entanglement entropy from ads/cft. *Physical review letters*, 96(18):181602, 2006.

[29] Shinsei Ryu and Tadashi Takayanagi. Aspects of holographic entanglement entropy. *Journal of High Energy Physics*, 2006(08):045–045, Aug 2006.

[30] Pasquale Calabrese and John Cardy. Entanglement entropy and quantum field theory. *Journal of Statistical Mechanics: Theory and Experiment*, 2004(06):P06002, Jun 2004.

[31] Christoph Holzhey, Finn Larsen, and Frank Wilczek. Geometric and renormalized entropy in conformal field theory. *Nuclear Physics B*, 424(3):443–467, Aug 1994.

[32] G. Vidal, J. I. Latorre, E. Rico, and A. Kitaev. Entanglement in quantum critical phenomena. *Physical Review Letters*, 90(22), Jun 2003.

[33] J. D. Brown and Marc Henneaux. Central charges in the canonical realization of asymptotic symmetries: an example from three-dimensional gravity. *Communications in Mathematical Physics*, 104(2):207 – 226, 1986.

[34] Leonard Susskind and Edward Witten. The holographic bound in anti-de sitter space. *arXiv preprint hep-th/9805114*, 1998.

[35] Kostas Skenderis. Lecture notes on holographic renormalization. *Classical and Quantum Gravity*, 19(22):5849–5876, Nov 2002.

[36] Veronika E Hubeny, Mukund Rangamani, and Tadashi Takayanagi. A covariant holographic entanglement entropy proposal. *Journal of High Energy Physics*, 2007(07):062–062, Jul 2007.

[37] Raphael Bousso. A covariant entropy conjecture. *Journal of High Energy Physics*, 1999(07):004–004, Jul 1999.

[38] Raphael Bousso. The holographic principle. *Reviews of Modern Physics*, 74(3):825–874, Aug 2002.

[39] Raphael Bousso and Netta Engelhardt. Generalized second law for cosmology. *Physical Review D*, 93(2), Jan 2016.

[40] Aron C Wall. Ten proofs of the generalized second law. *Journal of High Energy Physics*, 2009(06):021–021, Jun 2009.

[41] Tatsuma Nishioka, Shinsei Ryu, and Tadashi Takayanagi. Holographic entanglement entropy: an overview. *Journal of Physics A: Mathematical and Theoretical*, 42(50):504008, Dec 2009.

[42] Ahmed Almheiri, Raghu Mahajan, Juan Maldacena, and Ying Zhao. The page curve of hawking radiation from semiclassical geometry. *arXiv preprint arXiv:1908.10996*, 2019.

[43] Ahmed Almheiri, Netta Engelhardt, Donald Marolf, and Henry Maxfield. The entropy of bulk quantum fields and the entanglement wedge of an evaporating black hole. *Journal of High Energy Physics*, 2019(12), Dec 2019.

[44] Geoffrey Penington. Entanglement wedge reconstruction and the information paradox. *Journal of High Energy Physics*, 2020(9):1–84, 2020.

[45] Thomas Faulkner, Aitor Lewkowycz, and Juan Maldacena. Quantum corrections to holographic entanglement entropy. *Journal of High Energy Physics*, 2013(11), Nov 2013.

[46] Netta Engelhardt and Aron C. Wall. Quantum extremal surfaces: holographic entanglement entropy beyond the classical regime. *Journal of High Energy Physics*, 2015(1), Jan 2015.

[47] Chris Akers, Venkatesa Chandrasekaran, Stefan Leichenauer, Adam Levine, and Arvin Shahbazi Moghaddam. Quantum null energy condition, entanglement wedge nesting, and quantum focusing. *Physical Review D*, 101(2), Jan 2020.

[48] Stefan Leichenauer. The quantum focusing conjecture has not been violated. *arXiv preprint arXiv:1705.05469*, 2017.

[49] Chris Akers, Netta Engelhardt, Geoff Penington, and Mykhaylo Usatyuk. Quantum maximin surfaces, 2020.

[50] Leonard Susskind and John Uglum. Black hole entropy in canonical quantum gravity and superstring theory. *Physical Review D*, 50(4):2700–2711, Aug 1994.

[51] Aitor Lewkowycz and Juan Maldacena. Generalized gravitational entropy. *Journal of High Energy Physics*, 2013(8), Aug 2013.

[52] C. Brans and R. H. Dicke. Mach's principle and a relativistic theory of gravitation. *Phys. Rev.*, 124:925–935, Nov 1961.

[53] Xi Dong and Aitor Lewkowycz. Entropy, extremality, euclidean variations, and the equations of motion. *Journal of High Energy Physics*, 2018(1), Jan 2018.

[54] Aron C Wall. Maximin surfaces, and the strong subadditivity of the covariant holographic entanglement entropy. *Classical and Quantum Gravity*, 31(22):225007, Nov 2014.

[55] Lisa Randall and Raman Sundrum. An alternative to compactification. *Physical Review Letters*, 83(23):4690–4693, Dec 1999.

[56] J. Maldacena and L. Susskind. Cool horizons for entangled black holes. *Fortschritte der Physik*, 61(9):781–811, Aug 2013.

[57] Leonard Susskind and Larus Thorlacius. Gedanken experiments involving black holes. *Physical Review D*, 49(2):966, 1994.

[58] Leonard Susskind, Larus Thorlacius, and John Uglum. The stretched horizon and black hole complementarity. *Physical Review D*, 48(8):3743, 1993.

[59] Patrick Hayden and John Preskill. Black holes as mirrors: quantum information in random subsystems. *Journal of High Energy Physics*, 2007(09):120–120, Sep 2007.

[60] Matthew Headrick, Veronika E. Hubeny, Albion Lawrence, and Mukund Rangamani. Causality & holographic entanglement entropy. *Journal of High Energy Physics*, 2014(12), Dec 2014.

[61] Bartłomiej Czech, Joanna L Karczmarek, Fernando Nogueira, and Mark Van Raamsdonk. The gravity dual of a density matrix. *Classical and Quantum Gravity*, 29(15):155009, Jul 2012.

[62] John A Wheeler. Relativity, groups and topology. *edited by C. De-Witt and B. DeWitt, Gordon and Breach, New York*, 1964.

[63] John L. Cardy. Boundary conditions, fusion rules and the verlinde formula. *Nuclear Physics B*, 324(3):581–596, 1989.

[64] Jeff Steinhauer. Observation of quantum hawking radiation and its entanglement in an analogue black hole. *Nature Physics*, 12(10):959–965, 2016.

[65] Juan Ramon Munoz De Nova, Katrine Golubkov, Victor I Kolobov, and Jeff Steinhauer. Observation of thermal hawking radiation and its temperature in an analogue black hole. *Nature*, 569(7758):688–691, 2019.

[66] Victor I Kolobov, Katrine Golubkov, Juan Ramón Muñoz de Nova, and Jeff Steinhauer. Observation of stationary spontaneous hawking radiation and the time evolution of an analogue black hole. *Nature Physics*, 17(3):362–367, 2021.