

Master's thesis

# The Black Hole Information Paradox

Origins and Recent Insights

**William Eivik Olsen**

Theoretical Physics  
60 ECTS study points

Department of Physics  
Faculty of Mathematics and Natural Sciences

Autumn 2022





**William Eivik Olsen**

# The Black Hole Information Paradox

Origins and Recent Insights

Supervisor:  
Joakim Bergli



## Abstract

We give a thorough introduction to the black hole information paradox and discuss recent progress in solving it. Quantum field theory (QFT) in curved spacetime predicts that black holes emit exactly thermal radiation known as Hawking radiation. Since the radiation is exactly thermal, the entropy of Hawking radiation is predicted to increase monotonically as the black hole radiates. Meanwhile, the radiation carries away energy from the black hole, causing the thermodynamic entropy of the black hole, known as the Bekenstein-Hawking entropy, to decrease. At some point, roughly halfway into the evaporation, the QFT calculations predict that the entropy of Hawking radiation exceeds the Bekenstein-Hawking entropy. This should not be possible: For an outside observer, a black hole can be understood as a quantum system where the Bekenstein-Hawking entropy counts black hole microstates, as we argue by deriving the Bekenstein-Hawking entropy in the path-integral approach to quantum gravity. The entropy of Hawking radiation should therefore be a result of entanglement with the black hole quantum system, which implies that the entropy of Hawking radiation should be bounded by the Bekenstein-Hawking entropy, in conflict with the QFT result. This conflict is the black hole information paradox. If the entropy of Hawking radiation exceeds the Bekenstein-Hawking entropy, unitarity is broken and quantum information is lost. If the evaporation process is unitary, the entropy of Hawking radiation should follow a curve that initially increases, but eventually turns around and falls to zero, in agreement with the bound set by the Bekenstein-Hawking entropy. Recent progress in solving the paradox suggests that black hole evaporation is unitary. Using the path-integral approach to quantum gravity, the entropy of Hawking radiation was found to follow the curve one would expect of unitary evaporation. We give a general discussion of how this result was obtained.

## Abstract

# Contents

<b>1 Preliminaries</b>	<b>5</b>
1.1 Unitarity and Entropy . . . . .	5
1.2 General Relativity . . . . .	6
1.2.1 Causality . . . . .	6
1.2.2 Black Holes . . . . .	7
1.2.3 Black Hole Thermodynamics . . . . .	11
1.3 The Problem of Quantum Gravity . . . . .	12
<b>2 Hawking Radiation</b>	<b>15</b>
2.1 Quantum Field Theory in Curved Spacetime . . . . .	15
2.1.1 Application to Black Holes . . . . .	17
2.1.2 Black Hole Evaporation . . . . .	18
2.2 Breakdown of Predictability . . . . .	21
<b>3 Path Integrals</b>	<b>23</b>
3.1 Quantum Mechanics . . . . .	23
3.2 Quantum Field Theory . . . . .	24
3.3 The Replica Trick . . . . .	29
3.4 The Gravitational Path Integral . . . . .	32
3.4.1 The Gravitational Action . . . . .	34
3.4.2 The Gravitational Replica Trick . . . . .	35
<b>4 Path Integral Derivation of Hawking Radiation</b>	<b>37</b>
4.1 Rindler Space . . . . .	37
4.2 The Unruh Effect . . . . .	39
4.3 Hawking Radiation . . . . .	41
4.4 Thermofield Double . . . . .	42
4.5 The Euclidean Cigar . . . . .	45
<b>5 Why Information Loss is Paradoxical</b>	<b>47</b>
5.1 Path Integral Derivation of the Bekenstein-Hawking Entropy . . . . .	47
5.2 The Central Dogma . . . . .	49
5.3 The Information Paradox . . . . .	50
5.3.1 The Page Curve . . . . .	51
5.3.2 The Strength of the Information Paradox . . . . .	55

## Contents

<b>6</b>	<b>The Holographic Principle and the AdS/CFT Correspondence</b>	<b>57</b>
6.1	The Holographic Principle . . . . .	57
6.2	The AdS/CFT Correspondence . . . . .	58
6.3	AdS/CFT and the Information Paradox . . . . .	59
6.4	ER = EPR . . . . .	60
<b>7</b>	<b>Holographic Entanglement Entropy and Replica Wormholes</b>	<b>63</b>
7.1	Holographic Entanglement Entropy . . . . .	63
7.1.1	Entanglement Wedge Reconstruction . . . . .	64
7.1.2	Quantum Corrections . . . . .	64
7.2	Applying Quantum Extremal Surfaces to Black Holes . . . . .	66
7.3	Applying the Island Formula to Hawking Radiation . . . . .	67
7.4	Entanglement Wedges in Black Hole Evaporation . . . . .	69
7.5	The Replica Trick on the Cigar and a Proof of the RT Proposal . . . . .	70
7.6	Replica Wormholes . . . . .	73
<b>8</b>	<b>Summary and Discussion</b>	<b>79</b>
<b>A</b>	<b>Hypersurfaces</b>	<b>83</b>



# Acknowledgments

Det er mange som fortjener en takk for sitt bidrag til at jeg har fullført denne masteren. Først og fremst ønsker jeg å takke min veileder, Joakim, som overbeviste meg om å bytte til denne oppgaven, og som har vært til god hjelp hele veien. Jeg vil takke alle vennene jeg har fått i løpet av tiden min i Oslo, som har bidratt til at jeg har fått en fantastisk studietid. En stor takk går til alle på teoriseksjonen for det gode sosiale miljøet vi har der, hvor jeg særlig ønsker å gi en takk til Kevin, som har vært en god samtalepartner å diskutere informasjonsparadokset med. Jeg vil takke familien min, og til slutt, Alma, som har hørt litt mer enn hun kanskje ønsket om sorte hull.

## Acknowledgments

# Introduction

A black hole is a region of spacetime from which nothing can escape—at least as far as classical general relativity is concerned. In 1974, Stephen Hawking discovered that by including quantum effects in the spacetime background of a black hole, the black hole does emit particles [1, 2]. The emitted particles, referred to as *Hawking radiation*, follow a thermal spectrum of temperature  $T = 1/8\pi M$  in natural units, where  $M$  is the mass of the black hole. This means that the particles are in a mixed state. As it radiates, the black hole loses mass. After a long time—roughly  $10^{67}$  years for a solar-mass black hole—it has completely evaporated, and the universe is left containing only Hawking radiation. Assuming the black hole to have formed from a pure state, the process of black hole formation and evaporation is therefore pure-to-mixed and cannot be described as a unitary process. In this sense, black holes seem to destroy information about the state from which they formed. This led Hawking to suggest that gravity introduces a new level of uncertainty or randomness into physics over and above the uncertainty usually associated with quantum mechanics [3].

It is necessary to postulate unitary time evolution in order to make sense of the probabilistic interpretation in quantum mechanics. Therefore it seems that black hole evaporation and quantum mechanics are incompatible. This apparent tension is often referred to as the *black hole information paradox*. The general belief is that the information paradox is a result of our lacking understanding of quantum gravity. For this reason, the theoretical physics community has been searching for a solution to the paradox ever since the discovery of Hawking radiation, with the hope that it would lead to insights in quantum gravity.

The above description does not quite reflect the modern understanding of the information paradox, yet it is often described this way in less formal settings. There are many misconceptions surrounding the black hole information paradox, many of which can probably be traced back to the paradox being described in this way. Prominent critics of the paradox, including Penrose [4], Unruh and Wald [5, 6], correctly argue that unitary time evolution can only be expected in a globally hyperbolic spacetime. The spacetime of an evaporating black hole is not globally hyperbolic. Based on this argument, the authors claim that there is nothing paradoxical about black hole information loss. However, there *is* something deeply puzzling and paradoxical about black hole evaporation. The modern understanding is framed in terms of the entropy of the black hole and the radiation emitted from it.

Prior to the discovery of Hawking radiation, studies in classical general relativity had revealed that black holes have many properties reminiscent of thermodynamic systems. One of these properties is Hawking’s area theorem, which states that the surface area of a black hole cannot decrease [7]. There is an obvious similarity between this and the second law of thermodynamics. Bekenstein suggested that this is more than a similarity, and that black holes do have an entropy that is proportional to their surface

## Acknowledgments

area [8]. However, since a classical black hole cannot emit any particles, the second law of thermodynamics would be violated by, for example, a black hole immersed in black body radiation at a lower temperature than that of the black hole. The discovery of Hawking radiation fixed this issue and completed the description of black holes as thermodynamic systems. In addition, the value of the Hawking radiation temperature fixed the proportionality constant of the black hole entropy to  $1/4$ , giving the black hole entropy  $S_{\text{BH}} = A/4$ . This entropy is referred to as the *Bekenstein-Hawking entropy*.

If the Bekenstein-Hawking entropy is of similar origin as entropies in ordinary thermodynamics, it should be given by some counting of microstates in an underlying theory of quantum gravity. This does seem to be the case. The Bekenstein-Hawking entropy has been derived through several approaches to quantum gravity. Examples include Gibbons and Hawking [9], who did it in 1977 using the path-integral approach to quantum gravity, and Strominger and Vafa [10], who derived it in 1996 using string theory on a certain class of extremal Reissner-Nordström black holes.

This suggests that black holes can be described as an ordinary quantum system for an outside observer, with thermodynamic entropy given by the Bekenstein-Hawking entropy. If the black hole can be understood as a quantum system, the mixed state of Hawking radiation can be understood as a result of entanglement with the black hole. The thermodynamic entropy of the black hole must then be an upper limit for the entropy of Hawking radiation. As the black hole evaporates, this upper limit decreases. The problem is that the entropy of Hawking radiation will increase monotonically because the radiation is thermal. At some time called the *Page time*, the entropy of Hawking radiation is predicted to break the upper limit set by the Bekenstein-Hawking entropy. This conflict between the thermal nature of Hawking radiation, as predicted by quantum field theory in curved spacetime, and the quantum gravitational origin of the Bekenstein-Hawking entropy, is the modern understanding of the black hole information paradox.

The Page time is named after Don Page, who studied how the entropy of Hawking radiation should evolve if black holes are quantum systems to outside observers [11, 12]. He found that the entropy should increase at early times in agreement with the quantum field theory calculations of Hawking radiation. When the Page time is reached, the entropy should start to decrease, lying close to the maximal limit set by the Bekenstein-Hawking entropy. The resulting curve is called the *Page curve*.

Hawking was known as a defender of black hole information loss himself. That was until Juan Maldacena [13] discovered the AdS/CFT duality in 1997. AdS/CFT is a conjectured duality between a gravitational theory in Anti-de Sitter space (AdS) and a conformal field theory (CFT) that lives on the boundary of Anti-de Sitter space. It can be understood roughly as telling us that a system in quantum gravity can be described by a gravitational theory on the AdS side of the duality or equivalently by a conformal field theory on the CFT side of the duality. This means that a black hole in Anti-de Sitter space can be described by a CFT, which is only a type of quantum field theory and hence manifestly unitary. This famously convinced Hawking [14], and most of the community, that black hole evaporation must be a unitary process. In other words, there must be something wrong with the results one gets from the quantum field theory derivation of Hawking radiation. The entropy of Hawking radiation should somehow follow the Page curve.

The AdS/CFT duality opened for the recent years' significant progress in solving the paradox. One important concept introduced in AdS/CFT is *holographic entanglement entropy*, which relates the entropy on the CFT side of the duality to the entropy on the AdS side. In a recent series of papers [15, 16, 17, 18], a new formula for the

holographic entanglement entropy in quantum gravity was developed and applied to Hawking radiation, resulting in an entropy that followed the Page curve. The same result was later achieved without the AdS/CFT duality, instead using the path integral approach to quantum gravity [19, 20]. This was achieved by including topologically non-trivial geometries, referred to as *replica wormholes*, in the gravitational path integral.

The goal of this thesis is to give a thorough introduction to the black hole information paradox for a reader familiar with general relativity and quantum field theory, and to get an understanding of the recent gravitational path integral derivation of the Page curve. The gravitational path integral will be an important tool, both in performing Gibbons and Hawking's derivation of the Bekenstein-Hawking entropy so we can state the information paradox, and in studying the replica wormhole derivation of the Page curve. The thesis is structured as follows: In chapter 1, we set the stage by providing the necessary definitions of unitarity and entropy in quantum mechanics, in addition to the concepts of causality and black holes in general relativity. We have also included a discussion of the problem of quantum gravity. In chapter 2, we introduce quantum field theory in curved spacetime and give a general discussion of how the effect of Hawking radiation arises. We discuss the backreaction on the metric and describe the black hole evaporation process, and visit Hawking's original argument for information loss. In chapter 3, we give an introduction to path integrals and motivate the definition of the gravitational path integral. We also present how the entropy can be calculated in quantum field theory using the replica trick, which will later be needed for the replica wormhole calculations. In chapter 4, we show how path integrals can be used to show that black holes emit Hawking radiation. Next, we apply the gravitational path integral to the Schwarzschild black hole in chapter 5 and perform Gibbons and Hawking's derivation of the Bekenstein-Hawking entropy. Then we move on to give the modern statement of the information paradox and make a precise description of the Page curve. In chapter 6, we give a very brief introduction to the holographic principle and the AdS/CFT correspondence, and discuss their implications for the information paradox. Finally, in chapter 7, we discuss the recent progress on the paradox. We start the chapter by explaining the concepts of holographic entanglement entropy and state the conjectured formulas that include the effects of quantum gravity. We show qualitatively how these formulas give the Page curve for the entropy of both the black hole and the Hawking radiation. Then we perform the replica trick on the Schwarzschild black hole and use it to derive the Bekenstein-Hawking entropy again. We end the chapter by providing general arguments for the replica wormhole derivation of the Page curve, motivated by the replica trick on the Schwarzschild black hole. We summarize and discuss in chapter 8.

All calculations are done in units where  $\hbar = c = G = k_B = 1$ . We use the Lorentzian metric signature  $(-, +, +, +)$ .

## Acknowledgments

# Chapter 1

## Preliminaries

### 1.1 Unitarity and Entropy

The name ‘black hole information paradox’ is slightly misleading; the paradox is strictly speaking about *unitarity*. In order to conserve probability, a quantum mechanical state  $|\Psi(t)\rangle$  must evolve in time by a unitary operator  $\hat{U}(t, t_0)$ :

$$|\Psi(t)\rangle = \hat{U}(t, t_0) |\Psi(t_0)\rangle. \quad (1.1)$$

The requirement  $\hat{U}^\dagger = \hat{U}^{-1}$  ensures that an initially normalized state remains normalized at all times. Likewise, the time evolution of a density operator will be given by a unitary transformation

$$\rho(t) = \hat{U}(t, t_0) \rho(t_0) \hat{U}^\dagger(t, t_0). \quad (1.2)$$

A pure quantum state will remain pure under unitary time evolution. If a pure state somehow evolved into a mixed state, unitarity would be broken and we would say that ‘quantum information is lost’. It is in this sense we talk about ‘information’ in the black hole information paradox.

In order to describe what the black hole information paradox is about, we must first make a distinction between two different kinds of entropy that we will use. The simplest to describe is the von Neumann entropy, which is given in terms of the density matrix  $\rho$  by

$$S_{\text{vN}} = -\text{Tr}(\rho \log \rho). \quad (1.3)$$

The von Neumann entropy measures the purity of the system, with  $S_{\text{vN}} = 0$  for a pure state and  $S_{\text{vN}} > 0$  otherwise. An important property is that it is invariant under unitary time evolution.

If we have a quantum system composed of two parts  $A$  and  $B$ , we can define the density matrix of subsystem  $A$  by taking the partial trace with respect to subsystem  $B$ :

$$\rho_A = \text{Tr}_B(\rho). \quad (1.4)$$

The von Neumann entropy of subsystem  $A$  is then computed by inserting  $\rho_A$  in (1.3). Note that the system can be in a pure state with  $S_{\text{vN}}(A \cup B) = 0$  and still have entropy in the subsystems due to entanglement between them, giving them an equal entropy:  $S_{\text{vN}}(A) = S_{\text{vN}}(B)$ .

We often refer to the von Neumann entropy as the *fine-grained entropy*. This is in contrast with the other notion of entropy we will consider, which is the *coarse-grained entropy*. This entropy arises when we choose a subset of observables  $\{A_i\}$  ( $i = 1, \dots, N$ )

to describe the system of density matrix  $\rho$ . That is, the system is described by specifying

$$\langle A_i \rangle \equiv \text{Tr}(A_i \rho), \quad i = 1, \dots, N. \quad (1.5)$$

We define the coarse-grained entropy as [21, 22]

$$S_{\text{coarse}}(\{A_i\}, \rho) \equiv -\text{Tr}(\tilde{\rho} \log \tilde{\rho}), \quad (1.6)$$

where  $\tilde{\rho}$  is an effective density matrix that maximizes this quantity subject to the constraint

$$\langle A_i \rangle \equiv \text{Tr}(\tilde{\rho} A_i) = \text{Tr}(\rho A_i). \quad (1.7)$$

While the von Neumann entropy is invariant under unitary time evolution, the coarse-grained entropy is not. Instead, it satisfies the second law of thermodynamics,

$$\Delta S_{\text{coarse}} \geq 0. \quad (1.8)$$

A simple example of coarse-grained entropy is the ordinary entropy of thermodynamics, where the  $A_i$  are chosen to be a few observables such as energy and volume. Then the thermodynamic entropy is obtained by maximizing the von Neumann entropy among all states with that energy and volume.

A consequence of our definitions is that the fine-grained entropy must be smaller or equal to the coarse-grained entropy,

$$S_{\text{vN}}(\rho) \leq S_{\text{coarse}}(\rho). \quad (1.9)$$

Later we will see that black holes seem to break this inequality. This is what we will refer to as the black hole information paradox.

## 1.2 General Relativity

In this section, we will present concepts in general relativity that will be important throughout the thesis. We will do so following Carroll [23] unless otherwise stated.

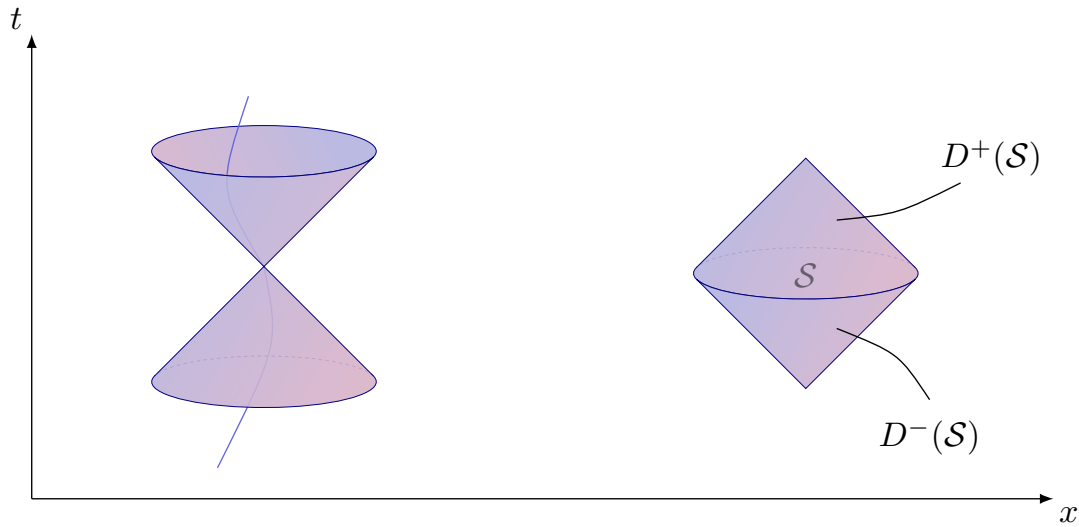
### 1.2.1 Causality

We are often interested in casting physical questions as initial-value problems: Given the state of a system at some moment of time, what will be the state some time later? We can give definite answers to such questions because of causality, the idea that future events can be understood as consequences of initial conditions plus the laws of physics. General relativity allows for spacetimes with very different causal structures from those of flat spacetime, introducing ways in which an initial-value formulation could break down. This will be important when we study black holes. Therefore, we give a brief introduction to how causality is understood in general relativity.

Assume that we have a fixed background metric of a manifold  $\mathcal{M}$  on which we evolve some matter fields. The guiding principle to discuss causality is the fact that any particle propagates at a speed limited by the speed of light. Since we use units in which  $c = 1$ , this can be stated as the restriction on a particle to propagate within its *lightcone* defined by  $dx/dt = \pm 1$ , as shown in figure 1.1. More formally, we say that the particle must follow a *causal curve*, defined to be a curve that is timelike or null everywhere.

A subset  $\mathcal{S} \subset \mathcal{M}$  is called *achronal* if no two points in  $\mathcal{S}$  are connected by a timelike curve. Given a closed achronal set  $\mathcal{S}$ , we define the *future domain of dependence* of





**Figure 1.1:** The lightcone (left) is defined by  $dx/dt = \pm 1$ . The trajectory of a particle is restricted to stay within its lightcone, as illustrated by the causal curve through the centre of the lightcone. The domain of dependence of an achronal region  $\mathcal{S}$  (right) is the part of the spacetime we can predict using initial conditions on  $\mathcal{S}$ .

$\mathcal{S}$ , denoted  $D^+(\mathcal{S})$ , as the set of all points  $p$  such that every past-moving causal curve through  $p$  must intersect  $\mathcal{S}$ . The *past domain of dependence*  $D^-(\mathcal{S})$  is defined accordingly by replacing future with past. We define the boundaries of  $D^+(\mathcal{S})$  and  $D^-(\mathcal{S})$  as the future and past *Cauchy horizons*  $H^+(\mathcal{S})$  and  $H^-(\mathcal{S})$ , respectively. Finally, we define the complete *domain of dependence* as  $D(\mathcal{S}) = D^+(\mathcal{S}) \cup D^-(\mathcal{S})$ . The domain of dependence for an achronal region  $\mathcal{S}$  is illustrated in figure 1.1. Given initial conditions on  $\mathcal{S}$ , we can predict what happens throughout the domain of dependence  $D(\mathcal{S})$ .

A closed achronal surface  $\mathcal{S}$  is said to be a *Cauchy surface* if the domain of dependence  $D(\mathcal{S})$  is the entire manifold. In that case, we can predict what happens throughout the entire manifold using initial conditions on  $\mathcal{S}$ , and we refer to the manifold as *globally hyperbolic*. A spacetime is not always hyperbolic: We will see that the spacetime of a black hole that forms from collapse and evaporates is an example where there is no Cauchy surface.

## 1.2.2 Black Holes

Birkhoff's theorem states that the Schwarzschild metric is the unique spherically symmetric solution for the vacuum field equations  $R_{\mu\nu} = 0$ . In spherical coordinates  $\{t, r, \theta, \phi\}$  it is given by

$$ds^2 = - \left(1 - \frac{r_s}{r}\right) dt^2 + \left(1 - \frac{r_s}{r}\right)^{-1} dr^2 + r^2 d\Omega^2, \quad (1.10)$$

where  $r_s = 2M$  is the Schwarzschild radius,  $M$  is the mass of the spherical source of energy-momentum, and  $d\Omega^2 = d\theta^2 + \sin^2 \theta d\phi^2$ . The metric may appear to be singular at  $r = 0$  and  $r = r_s$ . The origin is a true physical singularity; one can show that  $R^{\mu\nu\rho\sigma} R_{\mu\nu\rho\sigma} = 48G^2 M^2 / r^6$ , which diverges at  $r = 0$ . The Schwarzschild radius  $r = r_s$ , however, is not a physical singularity. It is only a result of our coordinate choices.

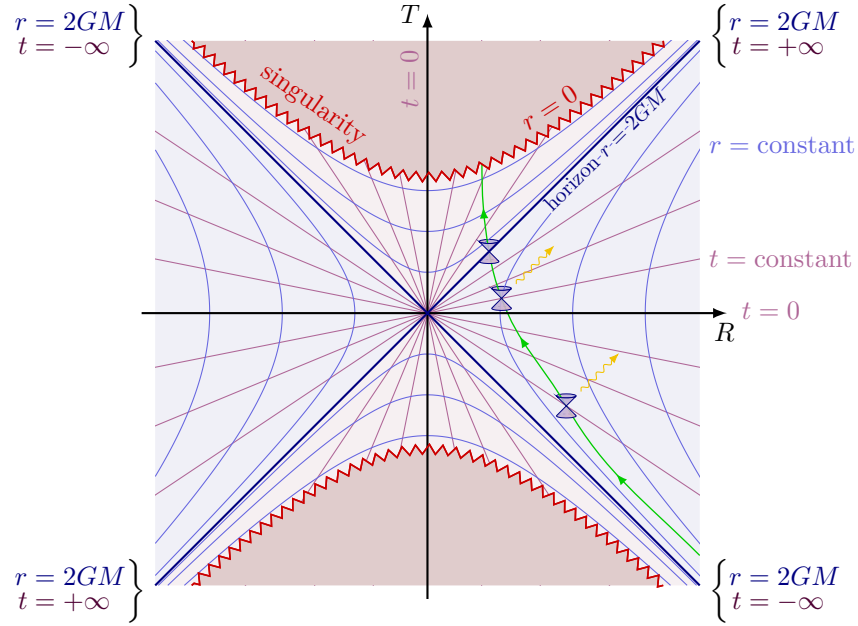


Figure 1.2: Kruskal diagram of the maximally extended Schwarzschild solution. Based on Tikz code by [24].

A more suitable coordinate system is the Kruskal coordinates  $\{T, R, \theta, \phi\}$ , in which the metric takes the form

$$ds^2 = \frac{32M^3}{r} e^{-r/r_s} (-dT^2 + dR^2) + r^2 d\Omega^2, \quad (1.11)$$

where  $r$  is defined implicitly from

$$T^2 - R^2 = \left(1 - \frac{r}{r_s}\right) e^{r/r_s}. \quad (1.12)$$

These coordinates have many nice properties. The radial null curves appear as they do in flat spacetime:

$$T = \pm R + \text{constant}. \quad (1.13)$$

There is an event horizon located at  $T = \pm R$ , which corresponds to  $r = r_s$  in spherical coordinates. More generally, we can consider the surfaces  $r = \text{constant}$ . From (1.12) we have that these satisfy

$$T^2 - R^2 = \text{constant}, \quad (1.14)$$

which means that they appear as hyperbolae in the  $R$ - $T$  plane. The surfaces of constant  $t$  are given by

$$\frac{T}{R} = \tanh\left(\frac{t}{2r_s}\right), \quad (1.15)$$

which define straight lines through the origin with slope  $\tanh(t/2r_s)$ . Note in particular that the surface  $t = \pm\infty$  represents the same surface as  $r = r_s$ .

We allow the  $T$  and  $R$  values to range over every value they can without hitting the singularity at  $r = 0$ . This corresponds to allowing the ranges  $R \in [-\infty, \infty]$ ,  $T^2 < R^2 + 1$ , giving the maximally extended Schwarzschild solution; the Kruskal coordinates cover the entire manifold described by the Schwarzschild solution. We can draw a spacetime

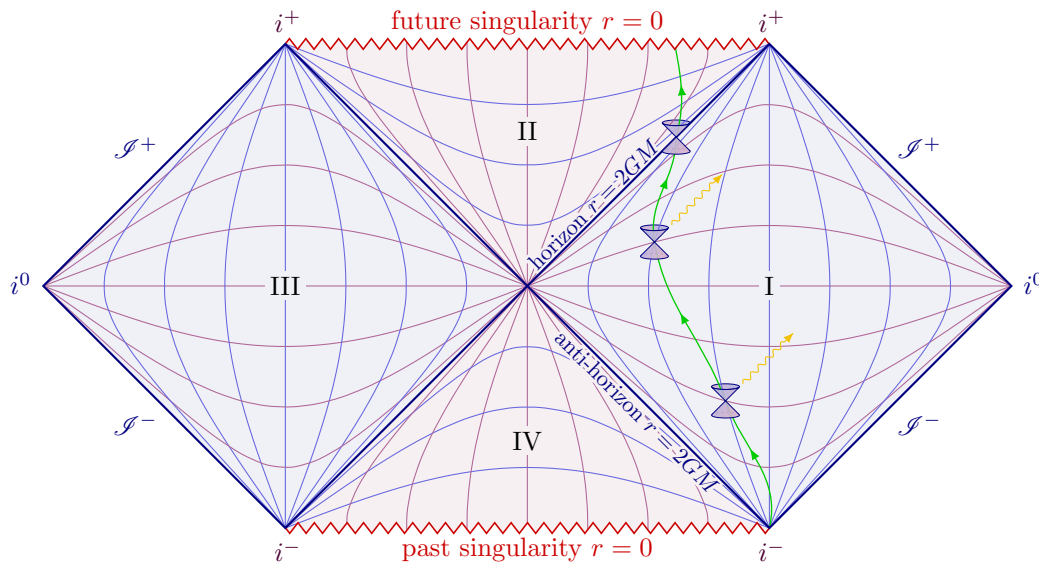


Figure 1.3: Penrose diagram of the maximally extended Schwarzschild solution. Based on Tikz code by [24].

diagram in the  $T$ - $R$  plane, suppressing the angular coordinates  $\theta$  and  $\phi$ , as shown in figure 1.2. This diagram is referred to as the *Kruskal diagram*.

We can get a very useful description of the causality of the spacetime if we collapse the coordinate ranges to a finite region and construct its *Penrose diagram*. Doing this on the Kruskal coordinates, one finds the Penrose diagram of figure 1.3.

Here  $i^-$  and  $i^+$  are past and future timelike infinity, respectively, and  $i^0$  is spatial infinity.  $\mathcal{I}^-$  and  $\mathcal{I}^+$  are past and future null infinity, respectively. The spacetime contains two asymptotically flat regions I and III. These two regions are connected by an *Einstein-Rosen bridge*, also referred to as a wormhole. Note that the wormhole is not traversable: Looking at the Penrose diagram, we see that a particle propagating on a timelike or spacelike trajectory in one of these two regions cannot reach the other region. Region II is a region from which no timelike or null geodesic can reach infinity. This is the black hole. Any particle in this region is destined to hit the future singularity. Region IV is in the causal past of both asymptotically flat regions. It is a region from which particles can reach the asymptotically flat regions, but which cannot be reached from the asymptotically flat regions. This is a white hole, the opposite of a black hole.

While the maximally extended Schwarzschild solution is a truly interesting spacetime, it is a highly idealized one that we do not expect to see in the real universe. It requires a spherically symmetric spacetime in which no energy-momentum is present and describes a black hole that has always and will always exist; for this reason it is often referred to as the ‘eternal black hole’.

The presence of energy-momentum will dramatically change the global picture. Birkhoff’s theorem implies only that the spacetime region outside the spherical source of energy-momentum is described by the Schwarzschild metric. The only requirement on the source of energy-momentum is that it is spherically symmetric. In particular, the source does not have to be static; it could be a collapsing star, as long as the collapse is spherically symmetric. The Penrose diagram of a black hole formed from stellar collapse is shown in figure 1.4. The future event horizon and the singularity are still present, but regions III and IV have been removed completely from the spacetime.

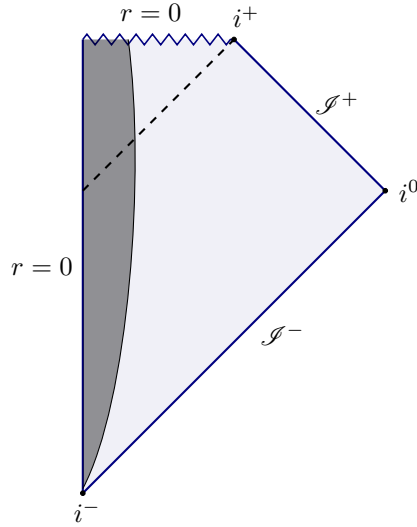


Figure 1.4: Penrose diagram of a Schwarzschild black hole that forms from collapse. The dark region is the collapsing star. The dashed line is the event horizon.

The singularity at  $r = 0$  is a place where the classical concept of spacetime as a manifold with a pseudo-Riemannian metric breaks down. Since all known laws of physics are formulated on a classical spacetime background, they will all break down at the singularity. One could hope that the singularity in the Schwarzschild solution is only a result of its spherical symmetry; this is in fact what many believed for a long time. That was until Penrose and Hawking proved their singularity theorems [25, 26, 27, 28]. The theorems can roughly be described as saying that given a high enough concentration of mass, a spacetime singularity is inevitable provided that general relativity is correct and that the energy-momentum tensor of matter satisfies a certain positive-definite inequality.

In classical general relativity, one can take solace in the idea that singularities are hidden behind event horizons and are therefore unavailable to an outside observer. This idea is referred to as the *cosmic censorship conjecture*. It has not been proven and is therefore only a conjecture. Several attempts have been made to disprove it, however, with none of them succeeding in the context of classical general relativity. This means that you should not fear encountering a *naked singularity* in your daily life; that is, a singularity from which signals can reach  $\mathcal{S}^+$ .

A consequence of cosmic censorship is that classical black holes cannot shrink. More precisely, the surface area  $A$  of the event horizon cannot decrease. This is Hawking's area theorem [7]:

$$\frac{dA}{dt} \geq 0. \quad (1.16)$$

The black holes we refer to in the singularity theorems and the area theorem are more general than that of the Schwarzschild solution. A general definition of a black hole in asymptotically flat spacetime is the following [29]:

A black hole  $\mathcal{B}$  in an asymptotically flat spacetime  $\mathcal{M}$  is the set of events that do not belong to the causal past of the future null infinity  $J^-(\mathcal{I}^+)$ , namely

$$\mathcal{B} = \mathcal{M} - J^-(\mathcal{I}^+) \neq \emptyset. \quad (1.17)$$

The event horizon is the boundary of the region  $\mathcal{B}$ .

One would naïvely expect that moving beyond spherical symmetry would allow for a rich variety of stationary black hole solutions. This turns out not be the case: All stationary black hole solutions can be described by a small number of parameters. The specific parameters depend on which matter fields we include in our theory; if electromagnetism is the only long-range nongravitational field, as it is in the standard model, we have a *no-hair theorem*:

Stationary, asymptotically flat black hole solutions to general relativity coupled to electromagnetism that are nonsingular outside the event horizon are fully characterized by the parameters of mass  $M$ , electric charge  $Q$  and angular momentum  $J$ .

Stationary black holes are the end states of gravitational collapse and can be seen as equilibrium states. For perturbations away from equilibrium, the change in mass is related to the change of area, electric charge and angular momentum by

$$dM = \frac{\kappa}{8\pi}dA + \Omega dJ + \Phi dQ, \quad (1.18)$$

where  $\kappa$  is the surface gravity,  $A$  the surface area,  $\Omega$  the surface angular velocity and  $\Phi$  the surface electric potential [30].

### 1.2.3 Black Hole Thermodynamics

Some of the black hole properties we have mentioned are very reminiscent of the laws of thermodynamics. This sparked the idea that black holes can be thought of as analogous to thermodynamic systems, and that stationary black holes are the equilibrium states of those systems.

Bardeen, Carter and Hawking formulated these properties as ‘the four laws of black hole mechanics’, which correspond to the four laws of thermodynamics [30]. They can be stated as follows:

- **The zeroth law** states that the horizon has a constant surface gravity  $\kappa$  for a stationary black hole.
- **The first law** states that for perturbations of stationary black holes, the change in mass satisfies (1.18).
- **The second law** states that the horizon area is a non-decreasing function of time, as stated by Hawking’s area theorem (1.16).
- **The third law** states that it is not possible, no matter how idealized, to reduce  $\kappa$  to zero by any finite sequence of operations.

Then the correspondence with the thermodynamic variables of energy  $E$ , entropy  $S$  and temperature  $T$  can be stated as

$$\begin{aligned} E &\leftrightarrow M \\ S &\leftrightarrow \lambda A \\ T &\leftrightarrow \kappa/8\pi\lambda, \end{aligned} \tag{1.19}$$

where  $\lambda$  is some unknown positive constant.

Bekenstein [8] went on to suggest that the correspondence between black holes and thermodynamics is more than an analogy: Black holes *are* thermodynamic systems in which the temperature and entropy are given by (1.19). The suggestion was motivated by considering the change in entropy as seen by an outside observer when matter with some entropy falls into a black hole: The outside observer would no longer have access to this matter, and the entropy of the universe would therefore appear to decrease, breaking the second law. That is, unless we make the identification  $S = \lambda A$  as the entropy of the black hole. Then we would say that the black hole contributes to the entropy, and we would have a *generalized entropy*:

$$S = \lambda A + S(\text{matter}), \tag{1.20}$$

where  $S(\text{matter})$  is the entropy of matter outside the black hole.

However, Hawking pointed out that the analogy breaks down when we try to bring the black hole into thermal contact with another system. By definition, nothing can escape from a black hole. Therefore we have no way of achieving heat flow from a black hole to another system. Thus we would be able to break the generalized second law by for example immersing the black hole in black body radiation with lower temperature than that of the black hole.

That is if we consider purely classical general relativity. Hawking [1, 2] wanted to confirm that black hole thermodynamics was only an analogy and that Bekenstein's interpretation was wrong even if quantum effects were included. But the result was the opposite: Hawking discovered that black holes *do* emit radiation, allowing black holes to be put in thermal contact with other systems. The radiation from black holes is known as *Hawking radiation*. Thus he ended up completing Bekenstein's interpretation of black holes as thermodynamic systems. Furthermore, he found that the emitted radiation is exactly thermal and has temperature

$$T = \frac{\kappa}{2\pi} = \frac{1}{8\pi M}. \tag{1.21}$$

This fixed the arbitrary constant in the entropy to  $\lambda = 1/4$ , leaving us with the generalized entropy

$$S = S_{\text{BH}} + S(\text{matter}), \tag{1.22}$$

where

$$S_{\text{BH}} = \frac{A}{4} \tag{1.23}$$

is the *Bekenstein-Hawking entropy*.

### 1.3 The Problem of Quantum Gravity

In this section, we discuss the problem of quantum gravity following Wald [31].

It is well established that general relativity is not the complete theory of gravity. The singularity theorems provide internal evidence that the theory is somehow incomplete. More suggestive is the fact that general relativity is a purely classical theory, while it is generally believed that physics is described on a fundamental level by the principles of quantum theory. It is therefore expected that general relativity is only a classical approximation to a truly fundamental theory of gravity.

Classical descriptions of ordinary matter are usually excellent approximations for describing phenomena which occur on macroscopic scales. This description breaks down on atomic and smaller scales. The scale at which this breakdown happens is determined by the masses and charges of the fundamental particles, and the two fundamental constants which enter the theory, namely the speed of light  $c$  and Planck's constant  $\hbar$ . Similarly, it is expected that in a quantum theory of gravity, the fundamental scale at which the classical description of general relativity breaks down is set by the fundamental constants of the theory, which are  $c$ ,  $\hbar$  and the gravitational constant  $G$ . Using these three constants, one can define a characteristic length scale  $l_P \equiv \sqrt{G\hbar/c^3} \sim 10^{-35}$  m referred to as the *Planck length*. It is expected that the classical concept of spacetime breaks down at length scales shorter than the Planck length, or time scales shorter than the *Planck time*  $t_P \equiv l_P/c \sim 10^{-43}$  s, equivalently. The characteristic energy at this scale is the *Planck energy*  $E_P \equiv \hbar/t_P \sim 10^{16}$  TeV.

Theorists have been searching for a quantum theory of gravity for a long time without convincing success. The difficulty compared to quantizing other classical field theories lies in the dual role played by the metric  $g_{\mu\nu}$  as both the quantity which describes the dynamical aspects of gravity and the quantity which describes the background spacetime structure. Therefore, it appears that in order to quantize the dynamical degrees of freedom of the gravitational field, one must also give a quantum mechanical description of spacetime. This has no analogue in other quantum field theories, which are formulated on a classically treated, fixed background spacetime.

One could imagine that these difficulties tell us that the principles of quantum theory do not apply to gravity, and general relativity *is* correct at the fundamental level. But Einstein's field equations

$$G_{\mu\nu} = 8\pi T_{\mu\nu} \tag{1.24}$$

tell us that the spacetime geometry is coupled to the matter sources. If the matter sources are described by a quantum theory, the geometry should also be quantized. To see this, let us assume that the opposite was true and we could describe the spacetime structure by a classical spacetime manifold  $\mathcal{M}$  with metric  $g_{\mu\nu}$ . The most natural candidate for a version of Einstein's equations where the matter content is described quantum mechanically would be the semiclassical Einstein equations

$$G_{\mu\nu} = 8\pi \langle \hat{T}_{\mu\nu} \rangle, \tag{1.25}$$

where  $\langle \hat{T}_{\mu\nu} \rangle$  denotes the expectation value of the stress-energy operator  $\hat{T}_{\mu\nu}$  in a given quantum state. Now imagine a state of matter where, with probability 1/2, the matter is located in a certain region  $A$  of spacetime and with probability 1/2 the matter is located in a region  $B$  disjoint from  $A$ . Equation (1.25) tells us that the gravitational field will behave as if half the matter is in  $A$  and the other half is in  $B$ . If we now measure the location of the matter, we will find it to be either in  $A$  or  $B$ . If (1.25) continues to hold after the measurement, the gravitational field must change in a discontinuous, acausal manner, in conflict with the principles of relativity. We could avoid these difficulties only by quantizing the gravitational field, such that it initially has probability 1/2 of having

the gravitational field configuration corresponding to matter in  $A$  and probability  $1/2$  of having the field corresponding to matter in  $B$ .

Although we lack a complete theory of quantum gravity, there are plenty of approaches under development, such as string theory and loop quantum gravity, in which there has been significant progress. However, all approaches share the common problem that we currently have no way of putting quantum gravity predictions to experimental tests: The Planck energy  $E_P \sim 10^{16}$  TeV is far beyond the maximum collision energy at the Large Hadron Collider, which is  $\sim 10$  TeV [32].

This makes the black hole information paradox very interesting for theorists wanting to study quantum gravity. It provides an explicit example where gravity and quantum mechanics conflict, thus acting as replacement for the lacking experimental data. The hope is that solving the paradox must provide insights into quantum gravity. This approach of finding apparent paradoxes between established theories and using them to develop new and more general theories has historically been successful. A good example is Einstein's discovery of the special theory of relativity as a solution to the conflict between the prediction of classical electromagnetism that the speed of light is the same in all inertial reference frames, and the Galilean principle of relative motion in Newtonian mechanics.

We will not go into the vast subjects of string theory and loop quantum gravity in this thesis. We will rather restrict our attention to the simpler *path integral approach to quantum gravity*, in which transition amplitudes are calculated directly from a sum over histories of possible spacetime geometries. This approach will be presented in chapter 3. The gravitational path integral has the advantage that the classical limit is easily obtained by extremizing the action; we shall apply it in this limit only. One can expect to obtain some effects of quantum gravity in the classical limit, just like we in some circumstances are able to reliably calculate the spontaneous creation of electron-positron pairs in a strong electric field using a classical treatment of the electromagnetic field. In fact, Hawking was able to derive the effect of Hawking radiation through the even simpler approach of *quantum field theory in curved spacetime*, in which quantum fields are treated on a fixed spacetime background. This approach will be presented in chapter 2.



## Chapter 2

# Hawking Radiation

Hawking incorporated quantum effects to the black hole spacetime by considering quantum field theory (QFT) in curved spacetime. His calculation was performed on the spacetime background of a Schwarzschild black hole formed from collapse. In this chapter, we will give a brief and general introduction to QFT in curved spacetime and use it to motivate the Hawking effect, following Carroll [23]. The actual calculations by Hawking are quite long and technical, so we will not present them here. For readers interested in a more detailed introduction, we suggest the original papers by Hawking [1, 2] and the textbook by Birrell and Davies [33].

### 2.1 Quantum Field Theory in Curved Spacetime

The idea behind QFT in curved spacetime is to treat quantum field theory on the fixed background of a general spacetime manifold. This means that one does not consider the influence of the quantum fields on the background. We can go from standard QFT to QFT in curved spacetime by the usual procedure in general relativity of minimal coupling. The Lagrangian density of a scalar field  $\phi(t, \vec{x})$  with mass  $m$  can then be written as<sup>1</sup>

$$\mathcal{L} = -\frac{1}{2}\sqrt{-g} \left( g^{\mu\nu} \nabla_\mu \phi \nabla_\nu \phi + m^2 \phi^2 \right). \quad (2.1)$$

The equation of motion is the Klein-Gordon equation:

$$\left( \square - m^2 \right) \phi = 0, \quad (2.2)$$

where  $\square \equiv g^{\mu\nu} \nabla_\mu \nabla_\nu$ . Now we want to write down the solution to this by writing down a complete, orthonormal set of modes in terms of which any solution may be expressed. To give a clear meaning to orthonormal modes we must define an inner product on the space of solutions of the Klein-Gordon equation. We define the inner product as

$$(\phi_1, \phi_2) = -i \int_\Sigma (\phi_1 \nabla_\mu \phi_2^* - \phi_2^* \nabla_\mu \phi_1) n^\mu \sqrt{\gamma} d^3x, \quad (2.3)$$

where  $\Sigma$  is a spacelike hypersurface with induced metric  $\gamma_{ij}$  and unit normal vector  $n^\mu$ . The inner product is independent of the choice of  $\Sigma$ .

---

<sup>1</sup>More generally, one can consider a coupling term between the scalar field and the curvature scalar  $R$ , given by  $-\xi R \phi^2$  for some constant  $\xi$ . We will not be needing this coupling for our purposes, so we simply set  $\xi = 0$ .

In ordinary QFT we would now simply go ahead and introduce a set of positive- and negative-frequency modes forming a complete basis for solutions to the Klein-Gordon equation. This time, however, there will not in general be any timelike Killing vector. This means that we will not in general be able to separate the solutions into time-dependent and space-dependent factors and cannot classify modes as positive- or negative-frequency. We can still find a set of basis modes, but there will be many such sets with no way of preferring one over another. The notion of a vacuum or a number operator will depend on which set we choose. This is a very important property of QFT in curved spacetime and, as we will see, is the key to the Hawking effect.

Consider a spatial hypersurface  $\mathcal{S}$  which is a Cauchy surface. Let  $f_i(x^\mu)$  denote an orthonormal complete set of positive frequency modes on  $\mathcal{S}$  that solve the Klein-Gordon equation and has conjugate modes  $f_i^*(x^\mu)$ ,

$$(f_i, f_j) = \delta_{ij}, \quad (f_i^*, f_j^*) = -\delta_{ij}, \quad (f_i, f_j^*) = 0. \quad (2.4)$$

We expand the fields in terms of the modes as

$$\phi(x^\mu) = \sum_i \left( \hat{a}_i f_i(x^\mu) + \hat{a}_i^\dagger f_i^*(x^\mu) \right), \quad (2.5)$$

and quantize the field by imposing commutation relations

$$[\hat{a}_i, \hat{a}_j] = 0, \quad [\hat{a}_i^\dagger, \hat{a}_j^\dagger] = 0, \quad [\hat{a}_i, \hat{a}_j^\dagger] = \delta_{ij}. \quad (2.6)$$

We define the vacuum state  $|0_f\rangle$  for this set of modes as the state that is annihilated by all the annihilation operators,

$$\hat{a}_i |0_f\rangle = 0 \quad \text{for all } i. \quad (2.7)$$

From the vacuum state, we can define an entire Fock basis  $\mathcal{H}(f_i)$  for the Hilbert space. We can create any state by applying the creation operator  $\hat{a}_i^\dagger$ ,

$$|n_i\rangle = \frac{1}{\sqrt{n_i!}} \left( \hat{a}_i^\dagger \right)^{n_i} |0_f\rangle, \quad (2.8)$$

and define the number operator for each mode as

$$\hat{n}_{fi} = \hat{a}_i^\dagger \hat{a}_i. \quad (2.9)$$

As we said, the choice of basis modes is not unique. We could have defined another set  $g_i(x^\mu)$  with creation and annihilation operators  $\hat{b}_i$  and  $\hat{b}_i^\dagger$  for which the exact same relations hold. In particular, this other set of modes will have its own vacuum state  $|0_g\rangle$  and number operator  $\hat{n}_{gi}$ . Since  $\{f_i\}$  serves as a complete set of solutions to the Klein-Gordon equation, we can write  $g_i$  as a linear combination of these and vice versa:

$$\begin{aligned} g_i &= \sum_j \left( \alpha_{ij} f_j + \beta_{ij} f_j^* \right), \\ f_i &= \sum_j \left( \alpha_{ji}^* g_j - \beta_{ji} g_j^* \right), \end{aligned} \quad (2.10)$$

where  $\alpha_{ij} = (g_i, f_j)$  and  $\beta_{ij} = -(g_i, f_j^*)$  are the *Bogolubov coefficients*. The Bogolubov coefficients can also be used to transform between the creation and annihilation

operators:

$$\begin{aligned}\hat{a}_i &= \sum_j \left( \alpha_{ji} \hat{b}_j + \beta_{ji}^* \hat{b}_j^\dagger \right), \\ \hat{b}_i &= \sum_j \left( \alpha_{ij}^* \hat{a}_j - \beta_{ij} \hat{a}_j^\dagger \right).\end{aligned}\tag{2.11}$$

Now imagine that we have a system in the  $f$ -vacuum  $|0\rangle_f$  in which no  $f$ -particles will be observed. We can find out how many particles will be observed by an observer using the  $g$ -modes by calculating the expectation value of the number operator  $\hat{n}_{gi}$ . Using the Bogolubov transformations to write the  $\hat{b}$  operators in terms of the  $\hat{a}$  operators, we find

$$\langle 0_f | \hat{n}_{gi} | 0_f \rangle = \sum_j |\beta_{ij}|^2.\tag{2.12}$$

This is generally non-zero. Thus we have that what looks like empty vacuum to one observer will contain particles to another observer. The vacuum state of each observer is defined according to how they divide between positive- and negative-frequency modes, which in turn is given by the proper time along their trajectory. An observer with proper time  $\tau$  will define positive frequency modes to obey

$$\frac{D}{d\tau} f_i = -i\omega f_i,\tag{2.13}$$

where  $\frac{D}{d\tau} = \frac{dx^\mu}{d\tau} \nabla_\mu$  is the directional covariant derivative. Using these modes we can calculate how many particles the observer will see.

### 2.1.1 Application to Black Holes

Let us now apply these ideas to the Schwarzschild black hole. Consider first the eternal black hole with the Penrose diagram of figure 1.3. We place two different observers in the spacetime. Let us name them Alice and Bob. Alice is a static observer outside the event horizon who measures time  $t$  and decomposes  $\phi$  in modes  $f_i, f_i^*$  with creation- and annihilation operators  $\hat{a}_i^\dagger, \hat{a}_i$ . Bob is an in-falling observer who measures time  $\tau$  and uses modes  $g_i, g_i^*$  with creation- and annihilation operators  $\hat{b}_i^\dagger, \hat{b}_i$ .

Now, similar to the procedure of regular QFT with interactions, we will define some *in*- and *out*-states. These states are defined to be in the vacuum state of a chosen observer in the remote past or future. We must choose a vacuum state such that the renormalized stress-energy tensor is well-defined and nonsingular. For the eternal Schwarzschild black hole the unique vacuum state that satisfies this is the *Hartle-Hawking vacuum*. The Hartle-Hawking vacuum is defined by letting Bob's in- and outgoing modes be in the vacuum. Therefore Bob will observe no particles. Alice, on the other hand, will observe particles. Doing the Bogolubov transformations, one would find that Alice placed at  $\mathcal{I}^+$  will observe a number density

$$\langle \hat{n}_\omega \rangle = \frac{\Gamma(\omega)}{e^{2\pi\omega/\kappa} \pm 1},\tag{2.14}$$

where  $\omega$  is the frequency of the  $f_i$  mode and  $\kappa$  is the surface gravity of the black hole.  $\Gamma(\omega)$  is a greybody factor that arises because of the backscattering of the particles that do not get past the potential well of the black hole. This is the exact number density one would expect from thermal radiation of temperature

$$T = \frac{\kappa}{2\pi} = \frac{1}{8\pi M}.\tag{2.15}$$

The Hartle-Hawking vacuum also contains an equal flux of thermal radiation from  $\mathcal{I}^-$  toward the black hole as seen by Alice. Thus the Hartle-Hawking vacuum describes a black hole in thermal equilibrium with its environment, as it should be if it is to describe an eternal black hole.

To describe the more realistic black hole of figure 1.4 that forms from collapse we can use the *Unruh vacuum*. This is defined by letting the ingoing modes be in Alice's vacuum state and letting the outgoing modes be in Bob's. In this state, Alice will observe outgoing particles with the thermal spectrum (2.14). However, there will not be an ingoing flux from  $\mathcal{I}^-$  as there is in the eternal black hole. In fact, this vacuum gives a singular energy-momentum tensor at the past horizon  $H^-$ . This does not concern us since we define it on the spacetime background of a black hole formed from collapse, in which there is no past horizon.

We could generalize the procedure to consider more general fields describing any particle species.<sup>2</sup> The result would be the same: The black hole emits thermal radiation, which in general is described by the thermal density matrix [3, 34]

$$\rho_{mn} = \delta_{mn} \frac{\Gamma^n (e^x \mp 1)^{\pm 1}}{(e^x \mp 1 \pm \Gamma)^{n \pm 1}}, \quad (2.16)$$

where the upper sign is for bosons, which can have any nonnegative integer number of quanta, and the lower sign is for fermions, which can only have  $n = 0$  or  $n = 1$ .  $\Gamma$  is the transmission coefficient of the hole to the particle of species  $s$  and rest mass  $\mu$  in a mode of energy  $x \equiv T\omega$ , helicity  $p$ , and angular momentum  $(l, m)$ . This is the Hawking effect: An observer at infinity will see the black hole emitting thermal radiation, with temperature given by (1.21) for the Schwarzschild black hole.

One can get a qualitative understanding of where Hawking radiation comes from by thinking of the pair production of virtual particles in the vacuum close to the event horizon, one of them with negative energy and the other with positive energy with respect to the timelike Killing vector at infinity. The negative energy particle is of course forbidden, but if it tunnels through the event horizon, it will reach a region where the Killing vector which represents time translations is spacelike. This means the particle can exist as a real particle with a timelike momentum vector even though its energy relative to infinity is negative. Beyond the event horizon, the particle will keep on falling inwards. The positive energy particle, on the other hand, can escape to infinity and will constitute what the outside observer sees as Hawking radiation. This radiation will be described by a density matrix because the particle pair is entangled. This picture is helpful, but it should not be taken too seriously. The real argument behind the Hawking effect is the calculations we have discussed.

## 2.1.2 Black Hole Evaporation

Hawking's calculations were performed on a fixed background metric. However, we have seen that the Hawking effect leads to a flux of thermal radiation being emitted from the black hole. Thus, by conservation of energy, we would expect the black hole to lose mass as it radiates, changing the background metric. This is called the *backreaction*. In order to calculate the backreaction exactly we would have to solve the semiclassical Einstein equation (1.25). We restate it here for convenience:

$$G_{\mu\nu} = 8\pi \langle T_{\mu\nu} \rangle, \quad (2.17)$$

<sup>2</sup>We could also allow the black hole to have angular momentum  $J$  and electric charge  $Q$  and would still get thermal radiation. Then  $x$  would also depend on  $J$  and  $Q$ .

where the expected energy tensor  $\langle T_{\mu\nu} \rangle$  gives a semiclassical description of the energy properties of the quantum fields. Unfortunately, this equation has never been solved for a number of reasons<sup>3</sup>. However, by considering the timelike Killing vector  $J^\mu$  of the Schwarzschild spacetime and making a few justified assumptions we can get a good approximation of how the spacetime will evolve.

The flux observed at infinity will be that of a black body that radiates at a temperature  $T$  inversely proportional to its mass  $M$ . From the Stefan-Boltzmann law we know that the flux emitted from a black body is proportional to  $AT^4$ , where the area  $A$  is proportional to  $M^2$  for a black hole. Thus we have that the flux is given by

$$F = \frac{\alpha}{M^2}, \quad (2.18)$$

where  $\alpha$  is some numerical coefficient that depends on which particle species can be emitted at a significant rate [36]. For emission of massless particles, which holds true for large black holes, the coefficient is  $\alpha \approx 3.7 \cdot 10^{-5}$  [37]. Since the spacetime is static outside of the collapsing matter, the energy current

$$J^\mu = K_\nu \langle T^{\mu\nu} \rangle \quad (2.19)$$

is conserved in that region. By conservation of  $J^\mu$ , a negative energy flux equal in magnitude to (2.18) must flow into the black hole. Thus it is clear that the backreaction causes the black hole to lose mass. For a black hole whose mass is much greater than the Planck mass, the backreaction effects will be locally small near the black hole horizon and in the entire exterior region [35]. Then the spacetime geometry should be accurately described by a locally Schwarzschild geometry in which the mass decreases slowly in time in agreement with (2.18). Thus, for a large black hole, we have

$$\frac{dM}{dt} = -F = -\frac{\alpha}{M^2}. \quad (2.20)$$

Solving (2.20) for a black hole of initial mass  $M_0$ , we obtain

$$M(t) = M_0 \left( 1 - \frac{t}{t_{\text{evap}}} \right)^{1/3}, \quad (2.21)$$

where

$$t_{\text{evap}} \equiv \frac{M_0^3}{3\alpha} \quad (2.22)$$

is the time it takes for the black hole to completely evaporate. Note that the approximation we used breaks down a Planck time before  $t_{\text{evap}}$ . At this point, we cannot continue to use the concept of a classical metric. However, the total mass left at this point is only a Planck mass. Thus, provided the black hole does not evolve into a negative mass naked singularity there is not much it can do except disappear altogether [2]. We are therefore left with a spacetime that can be described by the Penrose diagram of figure 2.1.

We should emphasize that, unlike the previous Penrose diagrams we have encountered, this one cannot be drawn from a set of coordinate transformations. It

---

<sup>3</sup>The equation is difficult both in principle and in practice. There are ambiguities in the definition of  $\langle T_{\mu\nu} \rangle$ . There is also a difficulty arising because  $\langle T_{\mu\nu} \rangle$  contains terms of fourth order in derivatives, whereas the classical Einstein equation is of second order. Finally, it is very difficult to calculate  $\langle T_{\mu\nu} \rangle$ , even for highly symmetric spacetimes. For more on these difficulties, see [35].

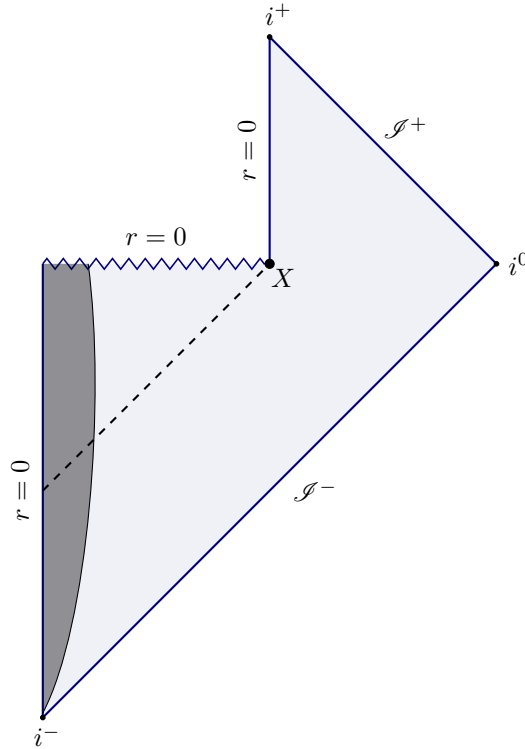


Figure 2.1: Penrose diagram of a Schwarzschild black hole that forms from collapse and evaporates.

is rather a hypothetical one in which one attaches a flat Minkowski spacetime to that of the Schwarzschild black hole from collapse which we saw in figure 1.4, motivated by our considerations of the backreaction. In order to find the correct Penrose diagram one would need a complete quantum gravitational description in which one could treat the evaporation. Thus we do not have much choice but accept this diagram.

Note that this spacetime has a naked singularity at the moment when the black hole completely evaporates, shown as  $X$  in the Penrose diagram. The naked singularity prevents us from defining a spacelike hypersurface that serves as a Cauchy surface. Therefore the spacetime of an evaporating black hole is *not* globally hyperbolic.

### Hilbert Space of a Partial Cauchy Surface

Recall that global hyperbolicity was a requirement if we were to make predictions over the entire spacetime. This was made explicit when we defined the modes  $f_i$  with which we solved the Klein-Gordon equation, as the modes were defined on a Cauchy surface. Nevertheless, it is still possible to do QFT on this spacetime background. Following [38], we can extend QFT in curved spacetime to *partial Cauchy surfaces*, which are defined as Cauchy surfaces for a given domain of dependence. An arbitrary spacelike hypersurface  $\Sigma$  is therefore a partial Cauchy surface for its domain of dependence  $D(\Sigma)$ . Then we can go ahead and repeat the exact same procedure of writing solutions to the Klein-Gordon equation with the only difference being that we now define modes  $f_i$  on the partial Cauchy surfaces. In this way, we can construct a Fock basis for the Hilbert space  $\mathcal{H}_f(\Sigma)$  on  $D(\Sigma)$ . Given any two sets of modes  $f, f'$ , complete on hypersurfaces  $\Sigma$ ,

$\Sigma'$ , respectively, we say the Hilbert spaces are *physically equivalent*,

$$\mathcal{H}_f(\Sigma) \sim \mathcal{H}_{f'}(\Sigma') \quad (2.23)$$

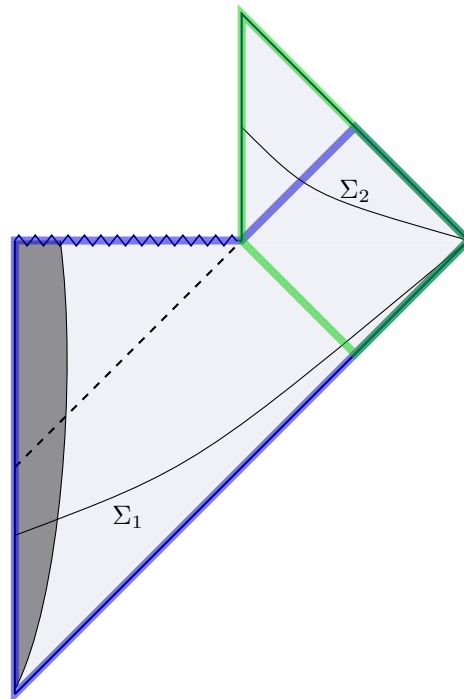
whenever the domains of dependence are equal,  $D(\Sigma) = D(\Sigma')$ . It is important to note that we expect unitary transformations between semiclassical Hilbert spaces *only* when they are physically equivalent.

## 2.2 Breakdown of Predictability

To summarize, Hawking's calculations show that a black hole emits thermal radiation described by the density matrix (2.16) for an observer at infinity, and it will evaporate as it does so. The spacetime left after evaporation is flat and contains only the thermal radiation. If the black hole formed from a pure state, the process of black hole evaporation is pure-to-mixed and therefore non-unitary. Thus we have that information is lost and an outside observer at late times cannot postdict the state from which the black hole formed. This led Hawking to suggest in his 1976 paper 'Breakdown of Predictability in Gravitational Collapse' [3] that gravity introduces a new level of uncertainty or randomness into physics over and above the uncertainty usually associated with quantum mechanics. Hawking concluded his paper by referring to the famous 'God does not play dice' quote by Einstein, to which he replies: 'God does not only play dice, He sometimes throws the dice where they cannot be seen.'

Hawking's argument that black hole evaporation is non-unitary is popularly referred to as the *black hole information paradox*. In such settings, the non-unitarity is often claimed to be paradoxical because 'quantum mechanics postulates unitary time evolution, so non-unitarity is incompatible with quantum mechanics'. This is not a valid argument against Hawking's claim: We have seen that unitary transformation between semiclassical Hilbert spaces is only expected when they are physically equivalent. The Hilbert spaces of partial Cauchy slices that extend from  $r = 0$  to infinity before and after evaporation are *not* equivalent, as can be seen in figure 2.2. In fact, Hawking never claimed there to be anything paradoxical about this non-unitarity in his paper.

However, we will see that there *is* something deeply puzzling and paradoxical about black hole information loss. This will become apparent when we try to answer where the Bekenstein-Hawking entropy comes from. The entropies we know from thermodynamics originate from statistical mechanics, in which the entropy counts the number of microstates for a given macrostate. Therefore, we expect that given a theory of quantum gravity we should be able to derive the Bekenstein-Hawking entropy through some counting of microstates. This turns out to be possible in several approaches to quantum gravity. We will present such a derivation using the *path integral approach* to quantum gravity, which we will develop in the next chapter.



**Figure 2.2:** Domains of dependence of partial Cauchy slices that extend from  $r = 0$  to infinity before ( $\Sigma_1$ ) and after ( $\Sigma_2$ ) evaporation.  $D(\Sigma_1)$  is marked in blue and  $D(\Sigma_2)$  in green. The domains of dependence are split by a Cauchy horizon due to the naked singularity present at the moment of complete evaporation. This makes their corresponding Hilbert spaces physically inequivalent and we therefore cannot expect unitary evolution.



## Chapter 3

# Path Integrals

In this chapter, we present useful tools from the path integral approach to quantum field theory and develop the gravitational path integral, which we will use to derive the Bekenstein-Hawking entropy in chapter 5.

### 3.1 Quantum Mechanics

The path integral formulation is an approach to quantum mechanics in which observables are expressed directly as integrals over classical trajectories between the initial and final configuration of a system. Before we define the path integral, we recap some basic quantum mechanics. States  $|\psi(t)\rangle$  in the Schrödinger picture evolve in time under the Schrödinger equation,

$$i \frac{d}{dt} |\psi(t)\rangle = \hat{H} |\psi(t)\rangle, \quad (3.1)$$

where  $\hat{H}$  is the Hamiltonian operator. Operators do not evolve in time in the Schrödinger picture. Therefore the eigenstate of the position operator  $\hat{x}$  is time independent:  $\hat{x}|x\rangle = x|x\rangle$ , where  $x$  is the eigenvalue. In the Heisenberg picture, it is the operators  $\hat{A}(t)$  that evolve in time under the Heisenberg equation

$$i \frac{d}{dt} \hat{A}(t) = [\hat{H}, \hat{A}(t)], \quad (3.2)$$

while the state vectors are time dependent. Therefore the eigenstate of the position operator  $\hat{x}$  is time dependent:  $\hat{x}(t)|x;t\rangle = x(t)|x;t\rangle$ , where  $x(t)$  is the eigenvalue. The position operator and eigenstate are related to those in the Schrödinger picture as

$$\hat{x}(t) = e^{i\hat{H}t} \hat{x} e^{-i\hat{H}t}, \quad |x;t\rangle = e^{i\hat{H}t} |x\rangle. \quad (3.3)$$

The transition amplitude from an initial state  $|x_i; t_i\rangle$  at time  $t_i$  to a final state at  $|x_f; t_f\rangle$  at time  $t_f$  is given by their overlap  $\langle x_f; t_f | x_i; t_i \rangle$  in the Heisenberg picture. If we translate to the Schrödinger picture, we get

$$\langle x_f; t_f | x_i; t_i \rangle = \langle x_f | e^{-i\hat{H}(t_f-t_i)} | x_i \rangle. \quad (3.4)$$

This transition amplitude can alternatively be calculated by the path integral, which results from splitting the time interval to many small intervals and inserting complete bases. The path integral is given by

$$\langle x_f | e^{-i\hat{H}(t_f-t_i)} | x_i \rangle = \int_{x(t_i)=x_i}^{x(t_f)=x_f} \mathcal{D}[x(t)] e^{iI_L[x(t)]}, \quad (3.5)$$

where  $I_L[x(t)] = \int_{t_i}^{t_f} dt L(x, \dot{x})$  is the Lorentzian action,  $L(x, \dot{x})$  is the Lagrangian of the given Hamiltonian and  $\mathcal{D}[x(t)]$  is the path integral measure [39]. The path integral is taken over all paths  $x(t)$  that satisfy the boundary conditions  $x(t_i) = x_i$ ,  $x(t_f) = x_f$ .

It is useful to analytically continue the Lorentzian time  $t$  to Euclidean time  $\tau$  by the Wick rotation  $\tau = it$ . This defines the *Euclidean path integral*

$$\langle x_f; \tau_f | x_i; \tau_i \rangle = \langle x_f | e^{-\hat{H}(\tau_f - \tau_i)} | x_i \rangle = \int_{x(\tau_i)=x_i}^{x(\tau_f)=x_f} \mathcal{D}[x(\tau)] e^{-I_E[x(\tau)]}, \quad (3.6)$$

where  $I_E[x(\tau)] = -iI_L[x(t)]$  is the Euclidean action. There is a close link between the Euclidean path integral and statistical mechanics. If we set  $\tau_f - \tau_i = \beta$ , the time evolution operator in Euclidean time becomes the thermal density matrix  $e^{-\beta \hat{H}}$ . If we also set  $x_f = x_i = x'$  and integrate over  $x'$ , we get the trace of the thermal density matrix, which is the thermal partition function  $Z(\beta)$ :

$$\int dx' \langle x' | e^{-\beta \hat{H}} | x' \rangle = \text{Tr} [e^{-\beta \hat{H}}] = Z(\beta). \quad (3.7)$$

Using (3.6) we find that the thermal partition function can be written as the Euclidean path integral

$$Z(\beta) = \int_{x(0)=x'}^{x(\beta)=x'} \mathcal{D}[x(\tau)] e^{-I_E[x(\tau)]} = \oint_{\beta} \mathcal{D}[x(\tau)] e^{-I_E[x(\tau)]}. \quad (3.8)$$

Thus we see that the inverse temperature  $\beta$  can be thought of as the Euclidean time period of the thermal partition function. In this way, we can use Euclidean path integrals to perform calculations in statistical mechanics, which we will find very useful throughout this thesis. We will in particular use it to calculate the entropy through

$$S = (1 - \beta \partial_{\beta}) \log Z(\beta). \quad (3.9)$$

## 3.2 Quantum Field Theory

In this section, we follow the lecture notes by Hartman [40] and the review article by Nishioka [41].

In quantum field theory, we define the path integral using field operators  $\hat{\phi}(t, \vec{x})$  in the Heisenberg picture. We define the eigenstate and eigenvalue of  $\hat{\phi}(t, \vec{x})$  by

$$\hat{\phi}(t, \vec{x}) |\phi(t, \vec{x})\rangle = \phi(t, \vec{x}) |\phi(t, \vec{x})\rangle. \quad (3.10)$$

The transition amplitude between field configurations  $|\phi_0(t_0, \vec{x})\rangle$  and  $|\phi_1(t_1, \vec{x})\rangle$  for  $t_1 > t_0$  is given by the path integral

$$\langle \phi_1(t_1, \vec{x}) | \phi_0(t_0, \vec{x}) \rangle = \langle \phi_1(\vec{x}) | e^{-i\hat{H}(t_1 - t_0)} | \phi_0(\vec{x}) \rangle = \int_{\phi(t_0)=\phi_0}^{\phi(t_1)=\phi_1} \mathcal{D}[\phi(t, \vec{x})] e^{iI_E[\phi(t)]}, \quad (3.11)$$

where  $|\phi_i(\vec{x})\rangle$  are eigenstates of the field operator  $\hat{\phi}(\vec{x})$  in the Schrödinger picture. We will skip writing the  $\vec{x}$  dependence unless it is needed. As before, we switch to Euclidean time and define the Euclidean path integral as

$$\langle \phi_1(\tau_1) | \phi_0(\tau_0) \rangle = \langle \phi_1 | e^{-\hat{H}(\tau_1 - \tau_0)} | \phi_0 \rangle = \int_{\phi(\tau_0)=\phi_0}^{\phi(\tau_1)=\phi_1} \mathcal{D}[\phi(\tau)] e^{-I_E[\phi(\tau)]}. \quad (3.12)$$

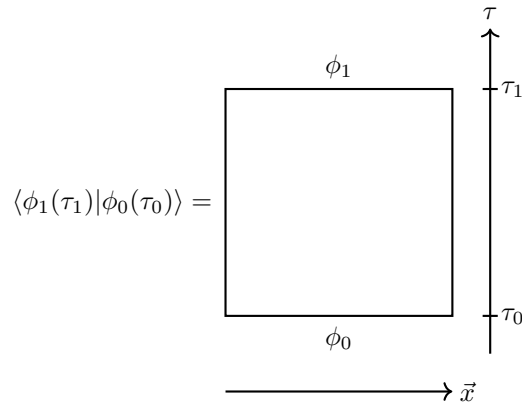
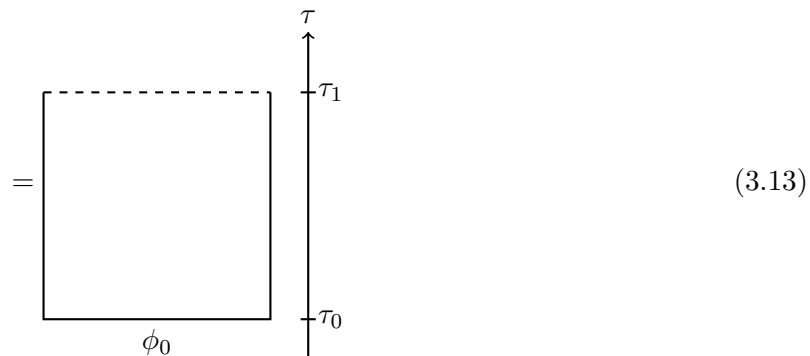


Figure 3.1: Drawing that represents (3.12).

We take the path integrals in QFT to be defined on any fixed background metric. It will be useful for us to represent the Euclidean path integral by drawing a diagram of the part of the spacetime we integrate over, as shown in figure 3.1. This is the integral over an infinite strip  $[\tau_0, \tau_1] \times R^{d-1}$ , where  $d$  is the number of spacetime dimensions.

The Euclidean path integral can be used to define a state  $|\Psi\rangle$ . This is achieved by only specifying one of the boundary conditions, leaving the other open:

$$|\Psi(\tau_1)\rangle = e^{-H(\tau_1-\tau_0)} |\phi_0\rangle = \int_{\phi(\tau_0)=\phi_0}^{\phi(\tau_1)=?} \mathcal{D}[\phi(\tau)] e^{-I_E[\phi(\tau)]}$$



We refer to such an open boundary as a *cut*.  $|\Psi\rangle$  is a functional that takes in a bra  $\langle\Omega|$  and returns a complex number  $\langle\Omega|\Psi\rangle$ . It follows that a bra  $\langle\Omega|$  should be given by a

path integral with a cut in the past:

$$\langle \Omega(\tau_0) | = \langle \phi_1 | e^{-H(\tau_1 - \tau_0)} = \int_{\phi(\tau_0)=?}^{\phi(\tau_1)=\phi_1} \mathcal{D}[\phi(\tau)] e^{-I_E[\phi(\tau)]}$$

(3.14)

More generally, any path integral with an open cut  $\Sigma$  defines a quantum state on  $\Sigma$ . We can also insert operators in the path integral to get a different state. We do this simply by adding the operator  $\hat{O}$  at the point  $(t, \vec{x})$  we want in the path integral. As an example, one can add two operators  $\hat{O}_1$  and  $\hat{O}_2$  at separate points to (3.13) to get a different state  $|\Psi'\rangle$  as follows:

$$|\Psi'\rangle = \int_{\phi(\tau_0)=\phi_0}^{\phi(\tau_3)=?} \mathcal{D}[\phi(\tau, \vec{x})] \hat{O}_1(\tau_1, \vec{x}_1) \hat{O}_2(\tau_2, \vec{x}_2) e^{-I_E[\phi(\tau, \vec{x})]}$$

(3.15)

The states we define using the Euclidean path integrals are still states in the Hilbert space of the Lorentzian theory. Consider the state  $|\Psi\rangle$  of (3.13). It is defined at a particular Lorentzian time  $t$ . We can evolve it in Lorentzian time by the Hamiltonian,

which can be drawn as a path integral over Lorentzian time:

$$\begin{aligned}
 |\Psi(t)\rangle &= e^{-i\hat{H}t} |\Psi\rangle \\
 &= \int_{\phi_0} \mathcal{D}\phi \exp(i\int_{t=0}^t \mathcal{L}(\phi, \dot{\phi}) dt) \\
 &= \int_{\phi_0} \mathcal{D}\phi \exp(i\int_{t=\tau=0}^t \mathcal{L}(\phi, \dot{\phi}) dt + \int_{\tau=0}^{\tau} \mathcal{L}(\phi, \dot{\phi}) d\tau)
 \end{aligned}
 \tag{3.16}$$

Note that we will always define the complex plane  $(t, \tau)$  such that  $t$  and  $\tau$  cross each other at  $t = \tau = 0$ , as shown in this path integral.

We can construct the ground state from the Euclidean path integral by integrating from  $\tau = -\infty$ . This can be seen by expanding the state in energy eigenstates

$$|\psi\rangle = \sum_n a_n |n\rangle, \quad \hat{H} |n\rangle = E_n |n\rangle.
 \tag{3.17}$$

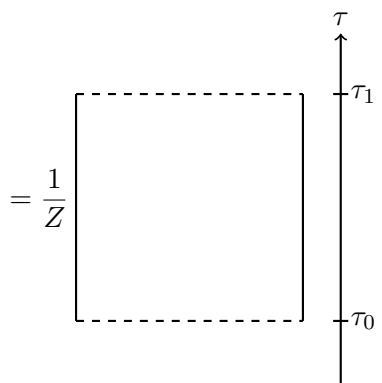
This gives

$$\lim_{\tau \rightarrow \infty} e^{-\hat{H}\tau} |\psi\rangle = \lim_{\tau \rightarrow \infty} \sum_n e^{-E_n \tau} |n\rangle \approx e^{-E_0 \tau} |0\rangle
 \tag{3.18}$$

because the term with the smallest energy  $E_0$  dominates in the limit. Thus we can write the ground state as the path integral

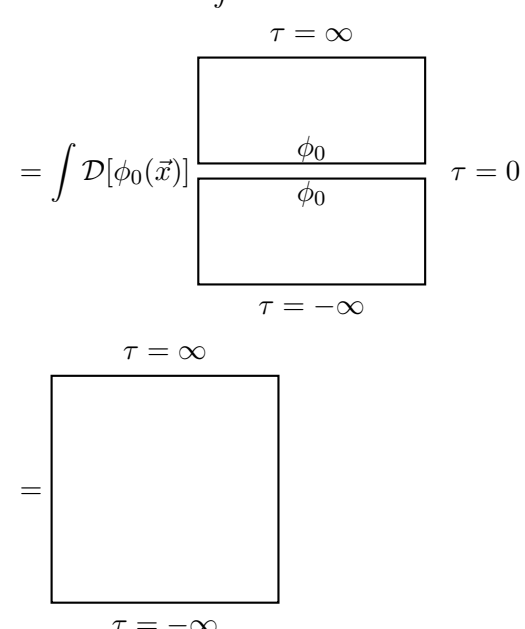
$$|0\rangle = \int_{\tau=-\infty}^{\phi(0)=?} \mathcal{D}[\phi(\tau)] e^{-I_E[\phi(\tau)]}
 \tag{3.19}$$

The path integral can also be used to construct density matrices. Since a density matrix is an operator that takes in a ket and a bra and returns a complex number, it should be defined by a path integral with two open cuts:

$$\rho = \frac{1}{Z} \int_{\phi(\tau_0)=?}^{\phi(\tau_1)=?} \mathcal{D}[\phi(\tau)] e^{-I_E[\phi(\tau)]}$$


$$= \frac{1}{Z} \tag{3.20}$$

Where the normalization factor  $Z = \text{Tr } \rho$  is the partition function of the density matrix. For the vacuum density matrix  $\rho_0 = |0\rangle \langle 0|$  it is given by

$$Z_0 = \text{Tr} [|0\rangle \langle 0|] = \int \mathcal{D}[\phi_0(\vec{x})] \langle \phi_0 | 0 \rangle \langle 0 | \phi_0 \rangle$$


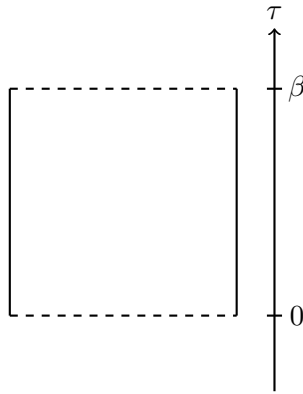
$$= \int \mathcal{D}[\phi_0(\vec{x})] \tag{3.21}$$

We see that taking the trace is equivalent to *gluing*  $|0\rangle$  and  $\langle 0|$  together; the integral over all possible boundary conditions  $\phi_0$  just says that the fields should be continuous at  $\tau = 0$  and therefore glues the open cuts together.

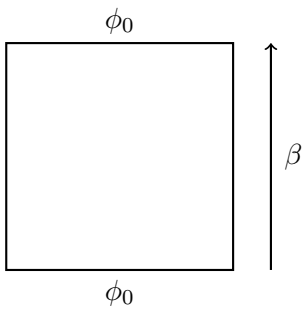
The thermal density matrix  $\rho_{\text{th}} = e^{-\beta \hat{H}}$  can be expressed as a path integral over a

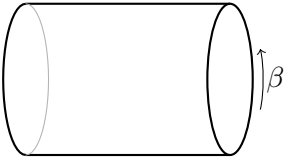
Euclidean time interval  $\beta$ :

$$\rho_{\text{th}} = \frac{1}{Z_{\text{th}}} \int_{\tau=0}^{\tau=\beta} \mathcal{D}[\phi(\tau)] e^{-I_E[\phi(\tau)]}$$

$$= \frac{1}{Z_{\text{th}}} \int_{\tau=0}^{\tau=\beta} \mathcal{D}[\phi(\tau)] e^{-I_E[\phi(\tau)]}$$

(3.22)

where its partition function  $Z_{\text{th}}$  is given by

$$Z_{\text{th}}(\beta) = \text{Tr} [e^{-\beta \hat{H}}] = \int \mathcal{D}[\phi_0(\tau)] \langle \phi_0 | e^{-\beta \hat{H}} | \phi_0 \rangle$$


$$= \int \mathcal{D}[\phi_0(\tau)] \langle \phi_0 | e^{-\beta \hat{H}} | \phi_0 \rangle$$


$$= \oint_{\beta} \mathcal{D}[\phi(\tau)] e^{-I_E[\phi(\tau)]}. \quad (3.23)$$

We see that calculating the thermal partition function is equivalent to glueing together the open cuts of the thermal density matrix, forming a cylinder periodic in Euclidean time. We also note that the Euclidean path integral for the thermal partition function in QFT is completely analogous to the quantum mechanical case (3.8).

### 3.3 The Replica Trick

The continuity of spacetime in QFT makes the von Neumann entropy more difficult to calculate than it is in quantum mechanics. In this section, we will present a way around this problem: The *replica trick*. We call it the replica trick because the trick is making  $n$  replicas of the system and calculating its Rényi entropy, given by

$$S_n(\rho) = -\frac{1}{n-1} \log \text{Tr} \rho^n. \quad (3.24)$$

The Rényi entropy is a generalization of the von Neumann entropy: If we analytically continue in  $n$  we can take the limit  $n \rightarrow 1$  to obtain the von Neumann entropy,

$$S_{\text{vN}}(\rho) = -\lim_{n \rightarrow 1} S_n(\rho) = -\lim_{n \rightarrow 1} \partial_n \log \text{Tr} \rho^n. \quad (3.25)$$

which is what we are going to do at the end of the calculation. We will follow the review article by Nishioka [41] closely.

First, we go back to our diagrams of path integrals in QFT and show how we can use path integrals to define the reduced density matrix. Consider the vacuum divided into regions  $A$  and  $B$  with total density matrix  $\rho_0 = |0\rangle\langle 0|$ . The reduced density matrix  $\rho_A$ , given by the partial trace over  $B$ , can be calculated by integrating the total density matrix over every state  $\phi^B(\vec{x})$ , where  $x \in B$ . This corresponds to writing down the path integral we had for the vacuum partition function  $Z_0$  (3.21) while leaving two open cuts in  $A$  at times  $\tau = 0^+$  and  $\tau = 0^-$ :

$$\begin{aligned} \rho_{0,A} &= \text{Tr}_B |0\rangle\langle 0| = \frac{1}{Z_0} \int \mathcal{D}[\phi_0(\vec{x} \in B)] \langle \phi_0 | 0 \rangle \langle 0 | \phi_0 \rangle \\ &= \frac{1}{Z_0} \int \mathcal{D}[\phi_0(\vec{x} \in B)] \begin{array}{c} \boxed{\begin{array}{cc} \phi_0 & \phi_0 \end{array}} \\ \text{-----} \\ \boxed{\begin{array}{cc} \phi_0 & \phi_0 \end{array}} \end{array} \begin{array}{l} \tau = 0^+ \\ \tau = 0^- \end{array} \\ \hspace{10em} \xrightarrow{\vec{x}} \\ \begin{array}{ccc} B & A & B \end{array} \\ \\ &= \frac{1}{Z_0} \begin{array}{c} \boxed{\text{-----}} \\ \tau = 0^+ \\ \tau = 0^- \end{array} \\ \hspace{10em} \xrightarrow{\vec{x}} \\ \begin{array}{ccc} B & A & B \end{array} \end{aligned} \quad (3.26)$$

The elements of  $\rho_{0,A}$  can be specified by inserting boundary conditions at the cuts:

$$\begin{aligned} [\rho_{0,A}]_{ij} &= \langle \phi_i^A | \rho_{0,A} | \phi_j^A \rangle \\ &= \frac{1}{Z_0} \begin{array}{c} \boxed{\begin{array}{c} \phi_i^A \\ \text{-----} \\ \phi_j^A \end{array}} \\ \tau = 0^+ \\ \tau = 0^- \end{array} \\ \hspace{10em} \xrightarrow{\vec{x}} \\ \begin{array}{ccc} B & A & B \end{array} \\ &= \frac{1}{Z_0} \int_{\tau=-\infty}^{\tau=\infty} \mathcal{D}[\phi(\tau, \vec{x})] e^{-I_E[\phi(\tau)]} \prod_{\vec{x} \in A} \delta(\phi(0^+, \vec{x}) - \phi_j^A(\vec{x})) \delta(\phi(0^-, \vec{x}) - \phi_i^A(\vec{x})). \end{aligned} \quad (3.27)$$



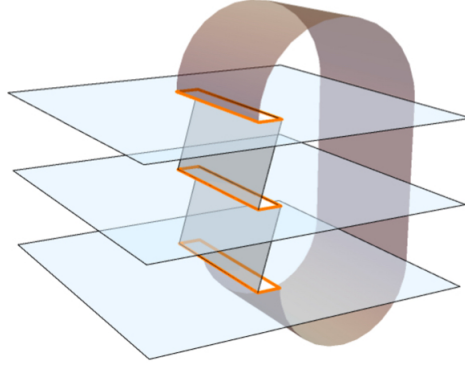


Figure 3.2: The replica manifold  $\mathcal{M}_n$  for  $n = 3$ . The cuts are glued together in a cyclic way. Adapted from [43].

We may then compute  $\text{Tr} \rho_{0,A}^n$  for any positive integer  $n$  by making  $n$  copies of the above, labelled by an integer  $k$  with  $1 \leq k \leq n$ , and gluing them together cyclically along the cuts so that  $\phi(\vec{x})'_k = \phi(\vec{x})_{k+1}$  (and  $\phi(\vec{x})'_n = \phi(\vec{x})_1$ ) for all  $\vec{x} \in A$ . Let us name this  $n$ -sheeted structure the *replica manifold*  $\mathcal{M}_n$  and the path integral on it  $Z_n$ . Then [42]

$$\begin{aligned}
 \text{Tr}_A(\rho_A^n) &= \frac{1}{(Z)^n} \begin{array}{c} \text{Diagram 1: Parallelogram with } \phi_2 \text{ above and } \phi_1 \text{ below a horizontal cut.} \\ \text{Diagram 2: Parallelogram with } \phi_3 \text{ above and } \phi_2 \text{ below a horizontal cut.} \\ \vdots \\ \text{Diagram n: Parallelogram with } \phi_1 \text{ above and } \phi_n \text{ below a horizontal cut.} \end{array} \\
 &= \frac{Z_n}{(Z)^n}. \tag{3.28}
 \end{aligned}$$

The cyclic glueing together is visualized more clearly in figure 3.2. Now we can insert in (3.25) to get

$$S_{\text{vN}}(\rho_A) = - \lim_{n \rightarrow 1} \partial_n (\log Z_n - n \log Z). \tag{3.29}$$

This is the replica trick; it allows us to calculate the von Neumann entropy directly from the partition functions  $Z$  and  $Z_n$ .

In most cases it is difficult to calculate  $Z_n$  on the replicated manifold  $\mathcal{M}_n$ . We can solve this by noticing that the cyclicity of the trace makes  $\mathcal{M}_n$  invariant under cyclic permutations of the replicas  $\mathcal{M}$ . We refer to this as the *replica symmetry*  $\mathbb{Z}_n$ . This means that we can construct a quotient space  $\hat{\mathcal{M}}_n \equiv \mathcal{M}_n / \mathbb{Z}_n$  that is topologically equivalent to  $\mathcal{M}$ . We can draw  $\hat{\mathcal{M}}_n$  as in figure 3.3. The quotient space will have  $n$  copies of the matter fields living on it.

In order to represent  $\mathcal{M}_n$  by  $\hat{\mathcal{M}}_n$  we must make some minor modifications. Assume that  $\mathcal{M}$  is a flat manifold that has matter fields living on it. Then  $\mathcal{M}_n$  will also be flat

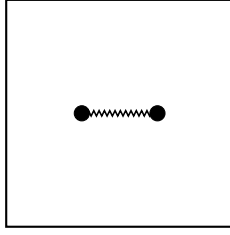


Figure 3.3: The quotient space  $\hat{\mathcal{M}} \equiv \mathcal{M}/\mathbb{Z}_n$  is topologically equivalent to  $\mathcal{M}$ . The zigzagged line represents the cuts that are glued together in  $\mathcal{M}_n$ . The black dots represent the insertion of twist operators.

everywhere *except* the endpoints of the regions that are glued together. At these points the curvature is infinite. This will affect the matter fields around these points. We can include this property in  $\hat{\mathcal{M}}_n$  by adding *twist fields* to the matter fields at the endpoints. We will not go into much detail on twist fields here. For more on them, see [44]. The important fact for our purposes is that the twist fields we are considering ensure that the  $n$  copies of the matter fields still have the symmetry under cyclic permutations. We can denote the cyclic permutation symmetry by  $\sigma$  and write it and its inverse as

$$\begin{aligned}\sigma &: i \mapsto i + 1 \pmod{n}, \\ \sigma^{-1} &: i + 1 \mapsto i \pmod{n}.\end{aligned}\tag{3.30}$$

Then we denote the twist fields associated with  $\sigma$  and  $\sigma^{-1}$  as  $\mathcal{T}(\tau, \vec{x})$  and  $\tilde{\mathcal{T}}(\tau, \vec{x})$ , respectively. Now let the spatial interval we considered in (3.28) be  $A = [\vec{u}, \vec{v}]$ . Then we have

$$Z_n \propto \langle \mathcal{T}(\tau = 0, \vec{u}) \tilde{\mathcal{T}}(\tau = 0, \vec{v}) \rangle_{\mathcal{L}^{(n)}, \hat{\mathcal{M}}_n},\tag{3.31}$$

where the subscript denotes that this correlator is evaluated in the  $n$ -copy model of the matter fields evaluated on  $\hat{\mathcal{M}}_n$ . If we allow the region  $A$  on which we calculate the reduced density matrix to be a set made of  $N$  disjoint intervals  $[\vec{u}_i, \vec{v}_i]$ , we have

$$Z_n \propto \langle \mathcal{T}(0, \vec{u}_1) \tilde{\mathcal{T}}(0, \vec{v}_1) \cdots \mathcal{T}(0, \vec{u}_N) \tilde{\mathcal{T}}(0, \vec{v}_N) \rangle_{\mathcal{L}^{(n)}, \hat{\mathcal{M}}_n}.\tag{3.32}$$

We will not do any explicit calculations where we use the twist operators and will therefore not use any of these equations directly. They will only be included when we later sketch the replica wormhole derivation of the entropy of Hawking radiation. The key fact about the twist operators we will need then is that they act as glueing together the matter fields on the disjoint intervals  $[u_i, v_j]$ . The manifold we consider then will have some *conical singularities* around which we also must add twist operators. The region around the conical singularities will therefore act as if there is an additional cut that must be glued together with the other cuts. The reader should not worry if this all seems rather vague at this point; it will become clearer when we get to the replica wormholes in chapter 7.

### 3.4 The Gravitational Path Integral

We will now present gravitational path integrals following Hawking [45] and Hartman [40]. In the QFT path integrals, we first fixed the spacetime manifold  $\mathcal{M}$ , then integrated

over all fields defined on  $\mathcal{M}$ . In quantum gravity, however, we must also integrate over the geometry itself. The gravitational path integral is

$$\int \mathcal{D}g \mathcal{D}\phi e^{-I_E[g,\phi]}, \quad (3.33)$$

where  $g$  denotes the metric  $g_{\mu\nu}$ ,  $\phi$  denotes the matter fields and  $I_E[g,\phi]$  is the Euclidean action, which is related to the Lorentzian action in the same way as before:  $I_E[g,\phi] = -iI_L[g,\phi]$ . The action will be defined in section 3.4.1.

The gravitational path integral must be assigned a boundary condition if it is going to have any meaning. In analogy to the QFT case, we define the thermal partition function  $Z(\beta)$  as the gravitational path integral on a Euclidean manifold with the boundary condition that Euclidean time has period  $\beta$  for an observer at infinity:

$$Z(\beta) = \int \mathcal{D}g \mathcal{D}\phi e^{-I_E[g,\phi]}. \quad (3.34)$$

In order to explicitly demand the boundary condition, we can specify the induced metric at spatial infinity as that of a flat cylinder:

$$h_{ab} dy^a dy^b = d\tau^2 + r_0^2 d\Omega^2 \text{ where } \tau + \beta \sim \tau, \text{ when } r_0 \rightarrow \infty. \quad (3.35)$$

The path integral is taken over all metrics of Euclidean signature  $(+,+,+,+)$  that induce the metric  $h_{ab}$  on  $\partial M$ . A useful visualization of the gravitational path integral for the partition function is shown in figure 3.4. The boundary condition is represented as the flat cylinder in the right part of the figure; the path integral is the sum of all the ways we can fill in the left part.

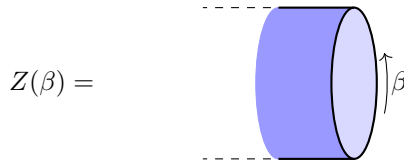


Figure 3.4: The gravitational path integral.

Note that in contrast to the path integral in QFT, this path integral cannot be derived. It is a postulate. Furthermore, the expression for the gravitational path integral is a formal expression that we cannot compute explicitly. However, we expect the dominant contributions to come from fields which are near a metric  $g_0$  and fields  $\phi_0$  which are an extremum of the action, which means that they solve the classical field equations. In fact, this must be the case if we are to recover classical general relativity in the limit of macroscopic systems. Let us then expand the action in a Taylor series about the background fields  $g_0, \phi_0$ ,

$$I_E[g,\phi] = I_E[g_0,\phi_0] + I_{E,2}[\tilde{g},\tilde{\phi}] + \text{higher-order terms}, \quad (3.36)$$

where  $g_{\mu\nu} = g_{0\mu\nu} + \tilde{g}_{\mu\nu}$ ,  $\phi = \phi_0 + \tilde{\phi}$  and  $I_{E,2}[\tilde{g},\tilde{\phi}]$  is quadratic in the perturbations  $\tilde{g}$  and  $\tilde{\phi}$ . If one ignores the higher-order terms, the path integral becomes

$$\log Z(\beta) = -I_E[g_0,\phi_0] + \log \int \mathcal{D}\tilde{g} \mathcal{D}\tilde{\phi} e^{-I_{E,2}[\tilde{g},\tilde{\phi}]}. \quad (3.37)$$

This is the *semiclassical approximation* of the gravitational path integral. We can regard the first term as the contribution of the background fields to  $\log Z$ . The second term is

called the one-loop term and represents the effect of quantum fluctuations around the background fields. In the case where all the background matter fields  $\phi_0$  are zero we can express the quadratic term as  $I_{E,2}[\tilde{g}, \tilde{\phi}] = I_{E,2}[\tilde{g}] + I_{E,2}[\tilde{\phi}]$  and

$$\log Z(\beta) = -I_E[g_0] + \log \int \mathcal{D}\phi e^{-I_{E,2}[\phi]} + \log \int \mathcal{D}\tilde{g} e^{-I_{E,2}[\tilde{g}]}. \quad (3.38)$$

Note that there can be several classical solutions  $g_0$  that satisfy the boundary condition (3.35). We refer to each of these solutions as *saddles*.

### 3.4.1 The Gravitational Action

In this section we follow Poisson [46], building on the formalism of hypersurfaces described in appendix A.

In order to compute the gravitational path integral we must first define the action. Let  $\mathcal{M}$  be an arbitrary region of the spacetime manifold, bounded by a closed hypersurface  $\partial\mathcal{M}$ . The Lorentzian action  $I_L[g, \phi]$  contains a contribution  $I_G[g]$  from the gravitational field and a contribution  $I_M[g, \phi]$  from the matter fields:

$$I_L[g, \phi] = I_G[g] + I_M[g, \phi]. \quad (3.39)$$

The matter action is given by

$$I_M[g, \phi] = \int_{\mathcal{M}} \mathcal{L}(\phi, \phi, \alpha; g_{\mu\nu}) \sqrt{-g} \, d^4x, \quad (3.40)$$

where  $\mathcal{L}$  is the Lagrangian density. The gravitational action is usually taken to be the Einstein-Hilbert action, which is defined as an integral over the Ricci scalar  $R$ :

$$I_H[g] = \frac{1}{16\pi} \int_{\mathcal{M}} R \sqrt{-g} \, d^4x. \quad (3.41)$$

There is an issue with this definition of the gravitational action: The Ricci scalar contains terms with second derivatives of the metric. The second derivatives show up as boundary terms in the variation of the action. This is a nontypical feature of field theories, which are usually formulated in terms of Lagrangians that involve dynamical variables and their first derivatives. We usually solve this issue by setting the boundary term to zero at infinity, which we are allowed to do when we are dealing with unbounded spacetime manifolds. However, when calculating the gravitational path integral (3.34) we must consider bounded spacetime manifolds in order to make sure the boundary condition (3.35) is satisfied. We must then deal with the issue by adding a term to the action that cancels the dependence on the second derivatives when we vary the action. The term we need to add is called the Gibbons-Hawking-York (GHY) boundary term, and can be written as

$$I_B[g] = \frac{1}{8\pi} \oint_{\partial\mathcal{M}} \varepsilon K |h|^{1/2} \, d^3y, \quad (3.42)$$

where  $K$  is the trace of the extrinsic curvature of  $\partial\mathcal{M}$ ,  $\varepsilon$  is equal to  $+1$  where  $\partial\mathcal{M}$  is timelike and  $-1$  where  $\partial\mathcal{M}$  is spacelike (it is assumed that  $\partial\mathcal{M}$  is nowhere null), and  $h$  is the determinant of the induced metric on  $\partial\mathcal{M}$ . Note that we use coordinates  $x^\alpha$  in  $\mathcal{M}$  and  $y^\alpha$  on  $\partial\mathcal{M}$ .

Adding the GHY boundary term is not enough to make the gravitational action well-defined: The term diverges for asymptotically flat spacetimes. We can solve this issue by subtracting a nondynamical term  $I_0$ , defined as

$$I_0 = \frac{1}{8\pi} \oint_{\partial\mathcal{M}} \varepsilon K_0 |h|^{1/2} \, d^3y, \quad (3.43)$$

where  $K_0$  is the extrinsic curvature of  $\partial\mathcal{M}$  embedded in flat spacetime. This makes the difference  $I_B - I_0$  finite, solving the issue.

Now we can add the terms together to form a well-defined gravitational action for bounded spacetime manifolds:

$$I_G[g] = I_H[g] + I_B[g] - I_0. \quad (3.44)$$

Finally, we can write the full Euclidean action  $I_E[g, \phi] = -iI_L[g, \phi]$  (using Euclidean time) as

$$I_E[g, \phi] = - \int_{\mathcal{M}} \left( \frac{R}{16\pi} + \mathcal{L} \right) \sqrt{g} \, d^4x - \frac{1}{8\pi} \oint_{\partial\mathcal{M}} \varepsilon (K - K_0) \sqrt{h} \, d^3y, \quad (3.45)$$

where  $g_{\mu\nu}$  and  $h_{ab}$  now have Euclidean signature.

### 3.4.2 The Gravitational Replica Trick

We can extend the replica trick to work for the gravitational path integral as well. This is achieved by defining the  $Z_n$  and  $Z$  in the replica trick (3.29) using the gravitational path integral. When we prepare a reduced density matrix using the gravitational path integral we will only specify the boundary condition; the rest is left for gravity to fill in. This will become clearer when we apply the gravitational replica trick to Hawking radiation in chapter 7.



## Chapter 4

# Path Integral Derivation of Hawking Radiation

A static observer outside a black hole observes Hawking radiation. By the equivalence principle, such an observer is locally equivalent to a constantly accelerating observer in flat spacetime. We therefore expect that an accelerating observer in flat spacetime should observe thermal radiation. This is indeed the case; it is called the *Unruh effect*. Since the effect involves an observer in flat spacetime, it is simpler to derive than the Hawking effect. Once the Unruh effect is found, one can use the equivalence principle to argue for the Hawking effect. We will therefore start by deriving the Unruh effect using the Euclidean path integral, then argue for the Hawking effect.

### 4.1 Rindler Space

In this section, we follow the discussion of Rindler space given in [47, 23].

Let us consider two-dimensional flat spacetime with the Minkowski metric

$$ds^2 = -dt^2 + dx^2. \quad (4.1)$$

Minkowski spacetime is invariant under the boost

$$\begin{aligned} t &\mapsto t \cosh \gamma + x \sinh \gamma, \\ x &\mapsto t \sinh \gamma + x \cosh \gamma, \end{aligned} \quad (4.2)$$

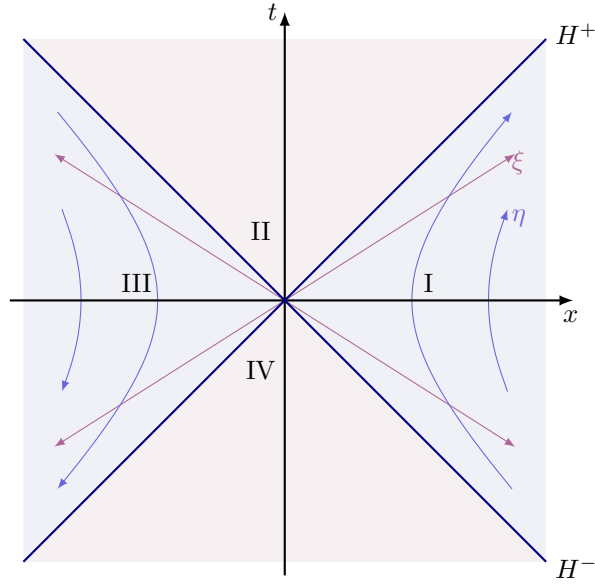
where  $\gamma$  is the boost parameter. The transformations are generated by the Killing vector  $x\partial_t + t\partial_x$ . The boost invariance motivates the introduction of new coordinates  $(\gamma, \rho)$  related to  $(t, x)$  by

$$\begin{aligned} t &= \rho \sinh \gamma, \\ x &= \rho \cosh \gamma, \end{aligned} \quad (4.3)$$

where  $\rho$  and  $\gamma$  take any real value. Then, the metric takes the form

$$ds^2 = -\rho^2 d\gamma^2 + d\rho^2, \quad (4.4)$$

This metric is referred to as the *Rindler metric*. The world lines with fixed values of  $\rho$  are the trajectories of the boost transformations. An observer following these worldlines experiences proper acceleration  $\alpha = 1/\rho$ . These coordinates cover only the



**Figure 4.1:** Rindler space. The Rindler observer is a constantly accelerating observer that follows the blue worldlines. The causal structure of Rindler space is similar to that of the maximally extended Schwarzschild spacetime, as can be seen in the Kruskal diagram of figure 1.2. The Rindler observer corresponds to an observer at fixed spatial coordinates in the Schwarzschild spacetime.

regions with  $x^2 > t^2$ . This defines two regions referred to as the left and right *Rindler wedges*, respectively, shown as regions I and III in figure 4.1. The regions not covered by the coordinates are II and IV. We will find it useful to make a further coordinate transformation to coordinates  $(\eta, \xi)$  related to  $(\gamma, \rho)$  by

$$\begin{aligned}\rho &= \frac{1}{a}e^{a\xi}, \\ x &= a\eta,\end{aligned}\tag{4.5}$$

where  $a$  is a positive constant and  $\eta$  and  $\xi$  can take any real value. We can also relate the new coordinates to  $(t, x)$  by

$$\begin{aligned}t &= \frac{1}{a}e^{a\xi} \sinh(a\eta), \\ x &= \frac{1}{a}e^{a\xi} \cosh(a\eta).\end{aligned}\tag{4.6}$$

Then, the metric takes the form

$$ds^2 = e^{2a\xi} (-d\eta^2 + d\xi^2).\tag{4.7}$$

These coordinates are useful because the world line with  $\xi = 0$  describes an observer with proper time  $\eta$  moving with a proper acceleration  $\alpha = a$ . Such an observer is called a *Rindler observer*. Since this metric is independent of  $\eta$ , it has a Killing vector  $\partial_\eta$ . If we express  $\partial_\eta$  in the Minkowski coordinates  $(t, x)$ , we find

$$\partial_\eta = \frac{\partial t}{\partial \eta} \partial_t + \frac{\partial x}{\partial \eta} \partial_x = a(x\partial_t + t\partial_x),\tag{4.8}$$



which we see is just the Killing vector associated with a boost in the  $x$ -direction times the acceleration  $a$ . This makes it clear that the Killing vector naturally extends throughout the spacetime. In region I it is timelike and future-directed, while in region III it is timelike and past-directed. In regions II and IV it is spacelike. Since the Killing vector in region III is past-directed, we can find describe it by coordinates  $(\tilde{\eta}, \tilde{\xi})$  where we flip the sign in (4.6),

$$\begin{aligned} t &= -\frac{1}{a}e^{a\tilde{\xi}} \sinh(a\tilde{\eta}), \\ x &= -\frac{1}{a}e^{a\tilde{\xi}} \cosh(a\tilde{\eta}). \end{aligned} \quad (4.9)$$

Looking at figure 4.1, we see that the causal structure of Rindler space is similar to the maximally extended Schwarzschild solution, as shown by the Kruskal diagram in figure 1.2. This fits in with our discussion of the equivalence principle. In particular, we see that the null line  $x = t$ , labelled  $H^+$ , is a future Cauchy horizon for a  $\eta = \text{constant}$  spacelike hypersurface in region I, and that the line  $x = -t$ , labelled by  $H^-$ , is a past Cauchy horizon. In fact,  $H^+$  and  $H^-$  are Killing horizons, defined as null hypersurfaces to which the Killing field is normal. The event horizons in the maximally extended Schwarzschild solution are also Killing horizons. Regions I and III then correspond to the two asymptotically flat regions in the maximally extended Schwarzschild solution.

## 4.2 The Unruh Effect

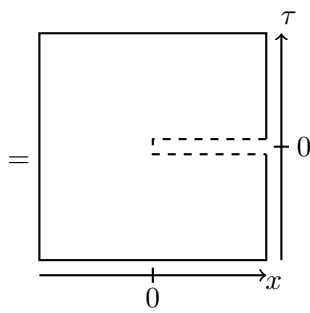
Having studied Rindler space, we are ready to derive the Unruh effect using path integrals. We do so following Hartman [40].

We begin by setting up the density matrix for the Minkowski vacuum  $|0\rangle$  at Euclidean time  $\tau = 0$  using the path integral

$$\begin{aligned} \rho &= |0\rangle \langle 0| \\ &= \begin{array}{c} \tau = \infty \\ \boxed{\phantom{\hspace{10em}}} \\ \text{---} \\ \tau = 0 \\ \text{---} \\ \boxed{\phantom{\hspace{10em}}} \\ \tau = -\infty \end{array} \quad (4.10) \end{aligned}$$

Next, we divide it into two regions  $B$ , for which  $x < 0$ , and  $A$ , for which  $x > 0$ . The density matrix for region  $A$  is obtained by tracing over region  $B$ , which is equivalent to

glueing the cuts together for  $x < 0$ :

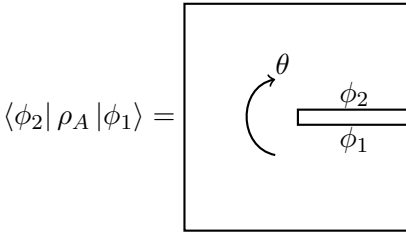
$$\rho_A = \text{Tr}_B \rho$$


$$(4.11)$$

Let us now calculate this path integral using the coordinates of the Rindler observer. Consider again equation (4.8) for the Killing vector  $\partial_\eta$ . Doing a Wick rotation transforms it to

$$\partial_\eta = ia(x\partial_\tau - \tau\partial_x) = ia\partial_\theta, \quad (4.12)$$

where  $\partial_\theta$  is the Killing vector that generates clockwise rotations in the  $(x, \tau)$  plane. Let us now specify boundary conditions  $\phi_1, \phi_2$  at the cuts and calculate  $\langle \phi_2 | \rho_A | \phi_1 \rangle$  using  $\theta$  as our time coordinate. This is equivalent to calculating  $\langle \phi_2 | \rho_A | \phi_1 \rangle$  by a rotation of  $2\pi$  from  $\phi_1$  to  $\phi_2$ :



$$\langle \phi_2 | \rho_A | \phi_1 \rangle = \quad (4.13)$$

We can construct a conserved quantity  $Q[\partial_\theta]$  related to the rotation Killing vector  $\partial_\theta$ . Then, by our usual rules for the path integral, this path integral must equal

$$\langle \phi_2 | \rho_A | \phi_1 \rangle = \langle \phi_2 | e^{-2\pi Q[\partial_\theta]} | \phi_1 \rangle, \quad (4.14)$$

where the conserved quantity  $Q[\partial_\theta]$  acts as the Hamiltonian that evolves the system a distance  $2\pi$  from the configuration  $|\phi_1\rangle$  to  $|\phi_2\rangle$ . Thus, we have

$$\rho_A = e^{-2\pi Q[\partial_\theta]}. \quad (4.15)$$

Furthermore, we can construct a conserved quantity  $Q[\partial_\eta]$  related to the timelike Killing vector  $\partial_\eta$ . By (4.12), we must have

$$Q[\partial_\theta] = -\frac{i}{a} Q[\partial_\eta]. \quad (4.16)$$

We can use a timelike Killing vector to define a conserved energy. If we denote the components of the Killing vector  $\partial_\eta$  by  $K_\mu$ , the conserved energy is given by an integral over a spacelike hypersurface  $\Sigma$ ,

$$Q[\partial_\eta] = \int_\Sigma K_\nu T^{\mu\nu} n_\mu \sqrt{h} \, d^3y, \quad (4.17)$$

where  $h_{ab}$  is the induced metric on  $\Sigma$  and  $n_\mu$  is the normal vector to  $\Sigma$ . Now, let  $\Sigma$  be the spacelike hypersurface defined by  $\tau = 0$ . The energy we get from (4.17) is then the

energy conserved in Euclidean time, with the normal vector  $n_\mu$  pointing in the positive Euclidean time direction. To get the physical energy conserved in Lorentzian time, we Wick rotate back. This also rotates the normal vector such that it points in the positive Lorentzian time direction. The new normal vector  $\tilde{n}_\mu$  is therefore related to the original normal vector  $n_\mu$  by  $n_\mu = i\tilde{n}_\mu$ . Thus we have

$$Q[\partial_\eta] = \int_\Sigma K_\nu T^{\mu\nu} n_\mu \sqrt{h} \, d^3y = \int_{\tilde{\Sigma}} K_\nu T^{\mu\nu} i\tilde{n}_\mu \sqrt{h} \, d^3y = iH, \quad (4.18)$$

where  $\tilde{\Sigma}$  is the spatial hypersurface defined by  $t = 0$ , and we have used that this coincides with the spatial hypersurface  $\Sigma$  defined by  $\tau = 0$ .  $H$  is the Hamiltonian, as it represents the conserved energy in Lorentzian time. Inserting in (4.15), we are left with

$$\rho_A = e^{-2\pi H/a}, \quad (4.19)$$

which we recognize as the thermal density matrix with temperature

$$T = \frac{a}{2\pi}. \quad (4.20)$$

This is the Unruh effect: A constantly accelerating observer detects thermal radiation at the temperature (4.20). In the language of QFT in curved spacetime, the effect is a result of the fact that the Rindler observer and the stationary Minkowski observer measure time differently, and therefore have different notions of vacuum.

### 4.3 Hawking Radiation

Now we are ready to argue for the Hawking effect. Consider the maximally extended Schwarzschild solution, described by the Kruskal diagram 1.2, with an observer at fixed coordinates just outside the event horizon in region I. By the equivalence principle, this observer should be locally equivalent to the Rindler observer in region I of Rindler space. To see this, let us take the Schwarzschild metric (1.10) and introduce a new radial coordinate  $\rho$  given by  $\rho^2 = 4r_s(r - r_s)$ . This transforms our metric into

$$ds^2 = - \left( 1 - \frac{4r_s^2}{\rho^2 + 4r_s^2} \right) dt^2 + \left( \frac{\rho^2}{4r_s^2} + 1 \right) d\rho^2 + r_s^2 \left( \frac{\rho^2}{4r_s^2} + 1 \right) d\Omega^2. \quad (4.21)$$

Expanding to second order in  $\rho$  near the horizon ( $\rho \rightarrow 0$ ), we find that the metric is

$$ds^2 \approx - \frac{\rho^2}{16M^2} dt^2 + d\rho^2 + r_s^2 d\Omega^2. \quad (4.22)$$

Let us also rescale the time coordinate to  $\gamma = t/4M$ , transforming our metric into

$$ds^2 \approx -\rho^2 d\gamma^2 + d\rho^2 + r_s^2 d\Omega^2, \quad (4.23)$$

where we see that the first two terms have exactly the form of the Rindler metric (4.4). Thus we have that the observer just outside the event horizon is a Rindler observer, and should detect Unruh radiation at the temperature (4.20).

Now, let us find the temperature observed at infinity. To do this, we can use the Tolman-Ehrenfest relation [48, 49], which states that

$$T \sqrt{g_{\mu\nu} K^\mu K^\nu} = \text{constant}, \quad (4.24)$$

where  $T$  is the locally measured temperature and  $K$  is the timelike Killing vector of the spacetime. For the Schwarzschild spacetime, this relation is then

$$T(r)\sqrt{-g_{tt}(r)} = \text{constant}, \quad (4.25)$$

where  $g_{tt} = -(1 - r_s/r)$  in the coordinates of (1.10). Now we can relate the temperature of an observer at infinity to that of an observer just outside the event horizon by

$$T(r = \infty) = T(r \sim r_s)\sqrt{-g_{tt}(r \sim r_s)}, \quad (4.26)$$

where we used that  $g_{tt}(r = \infty) = -1$ . For the observer just outside the horizon, we can use the near-horizon metric (4.22) in which  $g_{tt} = -\rho^2/(16M^2)$ , giving

$$T(r = \infty) = \frac{a}{2\pi} \frac{\rho}{4M} = \frac{1}{8\pi M}, \quad (4.27)$$

where we used that (4.5) gives  $\rho = 1/a$  for the Rindler observer, for which  $\xi = 0$ . Thus we have shown that an observer infinitely far from the eternal black hole observes Hawking radiation at temperature (1.21). Note that since we are considering the eternal black hole, we have radiation emitted from and entering the black hole; the vacuum state is the Hartle-Hawking state, as we described in chapter 2. As we also argued then, the vacuum state of a black hole formed from collapse is the Unruh state, in which there is only radiation emitted from the black hole and none entering. To get the density matrix at infinity, one must consider the scattering problem against the potential well surrounding the black hole. We will not show it here, but as stated in chapter 2, this gives a thermal density matrix with a greybody factor (2.16).

## 4.4 Thermofield Double

We will later find the concept of the *thermofield double* useful. The thermofield double is a state that purifies a thermal state  $|n\rangle$  by doubling the Hilbert space. It is given by

$$|\text{TFD}\rangle = \frac{1}{\sqrt{Z(\beta)}} \sum_n e^{-\beta E_n/2} |n\rangle_A |n\rangle_B, \quad (4.28)$$

where the energy eigenstates  $|n\rangle_A$  and  $|n\rangle_B$  live in separate Hilbert spaces  $\mathcal{H}_A$  and  $\mathcal{H}_B$ , with energy eigenvalue  $E_n$  [40]. The density matrix of the thermofield double state is given by

$$\rho = |\text{TFD}\rangle \langle \text{TFD}| \quad (4.29)$$

The reduced density matrix of system  $A$  is the thermal density matrix:

$$\rho_A = \text{Tr}_B \rho = e^{-\beta H_A}, \quad (4.30)$$

where  $H_A$  is the Hamiltonian of system  $A$ . Thus, if we consider only system  $A$ , the thermofield double state is indistinguishable from a thermal state. In the thermofield double state, the state in system  $A$  is thermal because it is entangled with system  $B$ .

The thermofield double can be prepared by the path integral

$$|\text{TFD}\rangle = \int_{\tau=0}^{\tau=\beta/2} \text{[Diagram]} \quad (4.31)$$

Note that this path integral looks exactly like one we would use to define a density matrix, since it has two cuts. This time, we will interpret the drawing differently. We interpret the lower cut as a cut for states in  $\mathcal{H}_A$  and the upper cut as a cut for states in  $\mathcal{H}_B$ , such that specifying boundary conditions  $\phi_1$  and  $\phi_2$  at the cuts is equivalent to multiplying by the bra  $\langle\phi_1|_A \langle\phi_2|_B$ .

We now make the claim that the Minkowski vacuum is equal to the thermofield double state, where the states  $|\phi_1\rangle_A$  and  $|\phi_2\rangle_B$  are states in the Hilbert spaces of Rindler observers in the left and right Rindler wedges. To show this, we prepare the vacuum state  $|0\rangle$  of the stationary Minkowski observer using a Euclidean path integral:

$$|0\rangle = \int_{\tau=-\infty}^{\tau=0} \text{[Diagram]} \quad (4.32)$$

Next, we divide the state into two regions  $A$  and  $B$ , corresponding to  $x > 0$  and  $x < 0$ , respectively, at Euclidean time  $\tau = 0$ . Then we specify  $\phi_1$  and  $\phi_2$  as boundary conditions for regions  $A$  and  $B$ , respectively, and calculate the transition amplitude

$$\langle\phi_1\phi_2|0\rangle = \int_{\tau=-\infty}^{\tau=0} \text{[Diagram]} \quad (4.33)$$

We repeat the trick we used in section 4.2, and switch time variable to that of the Rindler observer, corresponding to a clockwise rotation in the  $(x, \tau)$  plane. Then the

same transition amplitude is given by a clockwise rotation by  $\pi$  from  $|\phi_A\rangle$  to  $|\phi_B\rangle$ :

$$\begin{aligned}
 \langle \phi_1 \phi_2 | 0 \rangle &= \int_{\phi_1, \phi_2} \mathcal{D}\phi \exp(iS[\phi]) \\
 &= \int_{\phi_1, \phi_2} \mathcal{D}\phi \exp(iS[\phi]) \exp(i\theta) \\
 &= \langle \phi_1 |_A \langle \phi_2 |_B | \text{TFD} \rangle, \quad (4.34)
 \end{aligned}$$

where we in the final equality observed that this path integral is exactly the form of the thermofield double state with boundary conditions specified at the cuts. To find the inverse temperature  $\beta$  we must first convert back to the Rindler time  $\eta$ . By the same arguments as in section 4.2, this gives the inverse temperature

$$\beta = \frac{2\pi}{a}, \quad (4.35)$$

which is just the inverse of the Unruh temperature (4.20). While the result is the same as we found in section 4.2, we are now in a position to make some more interpretations. Consider figure 4.1, which describes Rindler space. The region  $A$  on which we calculated the density matrix for the Rindler observer is the surface  $t = 0, x > 0$ . The domain of dependence of this surface is the entire right Rindler wedge, shown as region I in the figure. This result tells us that we have a thermal state in the right Rindler wedge because it is entangled with the left Rindler wedge, and vice versa. In other words, we have that the vacuum of flat spacetime is an entangled state: As we discussed when explaining Hawking radiation in chapter 2, in the vacuum, pairs of entangled virtual particles are created and annihilated all the time.

This discussion can also be applied to the eternal black hole, since it has the same global structure as Rindler space: The thermal density matrix in region I can be interpreted as a result of entanglement between region I and III. The Hartle-Hawking vacuum is therefore equivalent to a thermofield double state, where the state  $|n\rangle_A$  and  $|n\rangle_B$  that enter are states in the Hilbert spaces of observers at fixed coordinates outside the black hole. The regions are entangled because they are part of the same spacetime; they are connected by a wormhole. As we have seen, the vacuum is an entangled state. Therefore, tracing out part of the vacuum leads to entanglement. This leads to some interesting interpretations when we later consider a black hole in anti-de Sitter space in chapter 6.

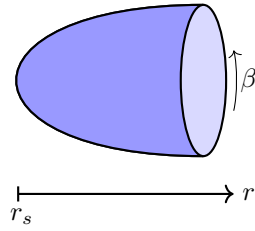


Figure 4.2: The Euclidean Schwarzschild spacetime, often referred to as the ‘cigar’.

## 4.5 The Euclidean Cigar

To complete the discussion, we will consider the Euclidean Schwarzschild metric. We turn the Schwarzschild metric (1.10) into a Euclidean metric by performing the Wick rotation  $\tau = it$ , giving the Euclidean Schwarzschild metric

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu = \left(1 - \frac{r_s}{r}\right) d\tau^2 + \left(1 - \frac{r_s}{r}\right)^{-1} dr^2 + r^2 d\Omega^2, \quad (4.36)$$

where we have defined  $x^\mu \equiv (\tau, r, \theta, \phi)$ . Note that  $r \geq r_s$  in the Euclidean Schwarzschild metric: For  $r < r_s$  the metric would have changed signature and would not represent a Euclidean manifold. Now let us do the same coordinate transformations we did to get the near-horizon metric (4.22), giving the Euclidean near-horizon metric

$$ds^2 \approx \frac{\rho^2}{16M^2} d\tau^2 + d\rho^2 + r_s^2 d\Omega^2. \quad (4.37)$$

We notice the resemblance between the first two terms and the metric of  $\mathbb{R}^2$  in Polar coordinates,  $ds^2 = dr^2 + r^2 d\theta^2$ . The angular coordinate in our Euclidean Schwarzschild metric is then  $\frac{\tau}{4M}$ . We are required to give the angular coordinate a period of  $2\pi$ . Choosing a different period  $2\pi - \alpha$  would correspond to cutting out a wedge of angle  $\alpha$  and glueing the remaining edges together, forming a cone. This would create a conical singularity at the horizon  $\rho = 0$ . However, we know that the Schwarzschild metric is perfectly well-behaved at the horizon. We are therefore forced to give  $\frac{\tau}{4M}$  a period of  $2\pi$ . Thus we identify  $\tau \sim \tau + \beta$ , where

$$\beta = 8\pi M. \quad (4.38)$$

We know that  $\beta$  is the inverse temperature of the quantum fields. Thus we have again obtained that the Schwarzschild black hole radiates particles with the temperature (1.21).

The Euclidean Schwarzschild spacetime can be drawn as in figure 4.2. We emphasize that the spacetime caps up smoothly at the event horizon  $r = r_s$ . The black hole interior and the singularity at  $r = 0$  has been completely excised, leaving an entirely nonsingular geometry. The funny shape of the Euclidean Schwarzschild spacetime has led to it being nicknamed the ‘cigar’.

It is useful to note that the spacelike surface  $\tau = 0$  from  $r_s$  to infinity corresponds to the surface  $T = 0$  from  $R = 0$  to infinity in the Kruskal diagram of the eternal black hole 1.2. Surprisingly, the Euclidean Schwarzschild metric also contains the other side of the eternal black hole on the other side of the cigar: The surface  $\tau = \beta/2$  from  $r_s$  to infinity corresponds to the surface  $T = 0$  from  $R = 0$  to negative infinity. To see this, consider Rindler space again, this time in the full complex time plane  $(t, \tau)$ , as shown in figure 4.3. We found that time evolution for the Rindler observer corresponds to

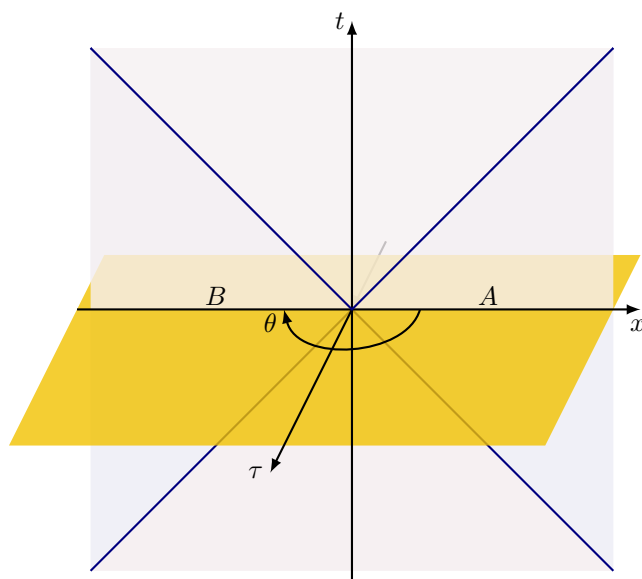


Figure 4.3: Complex Rindler spacetime. We can get from a point on  $A$  to a point in  $B$  by a rotation halfway around the Euclidean plane.

rotations in the Euclidean plane. Thus, using the Rindler coordinates, we can go from the right Rindler wedge to the left Rindler wedge by rotating by an angle of  $\theta = \pi$  in the Euclidean plane. Since a Rindler observer is locally equivalent to an observer at fixed spatial coordinates in Schwarzschild spacetime, this directly translates. Thus, rotating halfway around in Euclidean time takes us from a point in region I to a point in region III in the Schwarzschild spacetime. This discussion will be useful when we set up the density matrix at the end of chapter 7.



## Chapter 5

# Why Information Loss is Paradoxical

Statistical mechanics tells us that the thermodynamic entropy counts the number of microstates for a given macrostate. We may therefore expect that black holes have hidden microstates giving rise to the Bekenstein-Hawking entropy (1.23). In other words, if we knew the correct theory of quantum gravity, we should be able to derive the Bekenstein-Hawking entropy by counting the number of microstates that gives a black hole of a given mass, angular momentum and charge. In this section, we will review the evidence that this is indeed the case. We will focus on the path-integral approach to quantum gravity and show how this can be used to derive the Bekenstein-Hawking entropy.

### 5.1 Path Integral Derivation of the Bekenstein-Hawking Entropy

To derive the Bekenstein-Hawking entropy, we consider the gravitational path integral with no matter fields. We use the semiclassical approximation (3.38) and consider only the contribution of the background fields:

$$\log Z(\beta) \approx -I_E[g_0]. \quad (5.1)$$

We must then consider the contribution of all classical solutions that satisfy the boundary conditions (3.35) separately. Since we are considering a vacuum solution of the field equations we have  $R = \mathcal{L} = 0$ . The Euclidean action (3.45) is then simply

$$I_E = -\frac{1}{8\pi} \oint_{\partial\mathcal{M}} \varepsilon (K - K_0) \sqrt{h} d^3y. \quad (5.2)$$

In order to make sure the boundary conditions are satisfied we choose  $\partial M$  to be the large three-cylinder at  $r = r_0$ . This hypersurface corresponds to the coordinate restriction

$$\Psi(r) = r - r_0 = 0. \quad (5.3)$$

Note that  $\Psi$  is a timelike hypersurface.

There are two saddles that satisfy the boundary conditions: The Euclidean flat cylindrical spacetime

$$ds^2 = \eta_{\mu\nu} dx^\mu dx^\nu = d\tau^2 + dr^2 + r^2 d\Omega^2 \quad (5.4)$$

and the Euclidean Schwarzschild spacetime (4.36) with  $M = \frac{\beta}{8\pi}$ . Note that the Euclidean time coordinate in the flat cylinder (5.4) also has period  $\beta$ .

The induced metrics on  $\partial M$  from  $\eta_{\mu\nu}$  and  $g_{\mu\nu}$  are

$$ds^2 = u_{ab}dy^a dy^b = d\tau^2 + r_0^2 d\Omega^2, \quad (5.5)$$

$$ds^2 = v_{ab}dy^a dy^b = \left(1 - \frac{r_s}{r_0}\right) d\tau^2 + r_0^2 d\Omega^2, \quad (5.6)$$

respectively. We see that taking the limit  $r_0 \rightarrow \infty$ , both induced metrics satisfy the boundary condition (3.35).

Now we wish to find the action of both spacetimes. Since  $K_0$  is defined as the extrinsic curvature of flat spacetime, the action of flat spacetime is trivially  $I_E[\eta = 0]$ . Therefore the contribution of the flat cylinder to the logarithm of the partition function is zero. We still need to find  $K_0$  to compute the action from the Euclidean Schwarzschild spacetime. From (5.4) we see that  $\sqrt{\eta} = r^2 \sin \theta$ . Using (A.4), we find that the normal vector on  $\partial M$  is  $n^\mu = (0, 1, 0, 0)$ . Thus we get

$$\begin{aligned} K_0 &= \nabla_\mu n^\mu \\ &= \frac{1}{\sqrt{\eta}} \partial_\mu (\sqrt{\eta} n^\mu) \\ &= \frac{1}{r^2 \sin \theta} \partial_r (r^2 \sin \theta) \Big|_{r=r_0} \\ &= \frac{2}{r_0}. \end{aligned} \quad (5.7)$$

Next, we find the extrinsic curvature  $K$  of the Euclidean Schwarzschild spacetime. From (4.36) we see that  $\sqrt{g} = r^2 \sin \theta$ . From (A.4) we get that the normal vector on  $\partial M$  is  $n^\mu = \sqrt{1 - r_s/r_0} (0, 1, 0, 0)$ . This gives the extrinsic curvature

$$\begin{aligned} K &= \frac{1}{\sqrt{g}} \partial_\mu (\sqrt{g} n^\mu) \\ &= \frac{1}{r^2 \sin \theta} \partial_r \left( r^2 \sin \theta \sqrt{1 - \frac{r_s}{r}} \right) \Big|_{r=r_0} \\ &= \frac{2}{r_0} \sqrt{1 - \frac{r_s}{r_0}} + \frac{1}{2} \frac{r_s}{r_0^2} \frac{1}{\sqrt{1 - r_s/r_0}}. \end{aligned} \quad (5.8)$$

Inserting in (5.2), we get

$$\begin{aligned} I_E[g] &= -\frac{1}{8\pi} \int_{\partial M} \varepsilon (K - K_0) \sqrt{\gamma} d^3z \\ &= -\frac{1}{8\pi} \int_0^\beta d\tau \int_0^\pi d\theta \int_0^{2\pi} d\phi r_0^2 \sin \theta \left[ \frac{2}{r_0} \sqrt{1 - \frac{r_s}{r_0}} \right. \\ &\quad \left. + \frac{1}{2} \frac{r_s}{r_0^2} \frac{1}{\sqrt{1 - r_s/r_0}} - \frac{2}{r_0} \right] \\ &= \beta \left[ r_0 - r_0 \sqrt{1 - \frac{2M}{r_0}} - \frac{M}{2} \frac{1}{\sqrt{1 - 2M/r_0}} \right], \end{aligned} \quad (5.9)$$

where we have inserted  $r_s = 2M$ . Now we expand the square root of the second term in

the Taylor series  $\sqrt{1-x} = 1 - \frac{x}{2} + \mathcal{O}(x^2)$  to get

$$\begin{aligned} I_E[g] &= \frac{\beta}{2} \left[ r_0 - r_0 \left( 1 - \frac{M}{r_0} + \mathcal{O}(r_0^{-2}) \right) - \frac{M}{2} \frac{1}{\sqrt{1-2M/r_0}} \right] \\ &= \beta \left[ M - \frac{M}{2} \frac{1}{\sqrt{1-2M/r_0}} - \mathcal{O}(r_0^{-1}) \right]. \end{aligned} \quad (5.10)$$

Taking the limit  $r_0 \rightarrow \infty$ , we arrive at

$$I_E[g] = \frac{\beta M}{2} = \frac{\beta^2}{16\pi}. \quad (5.11)$$

Inserting in (5.1), we get

$$Z(\beta) \approx e^{-\beta^2/16\pi}. \quad (5.12)$$

Finally, we obtain the entropy of the black hole through (3.9):

$$S_{\text{BH}} = \frac{\beta^2}{16\pi} = 4\pi M^2 = \frac{A}{4}, \quad (5.13)$$

in agreement with the Bekenstein-Hawking formula (1.23).

Let us make a few remarks about this derivation. First of all, it is important to discuss why we interpret the entropy we got out of this derivation as that of the black hole itself. This is because we chose a classical solution with no matter fields—neither classical nor perturbations—and calculated its contribution to the partition function. Any entropy present must therefore be a consequence of the geometry itself. Second, we should discuss what kind of entropy we have found. Since it is the entropy of a classical black hole solution, it is the entropy of a geometry that can be described by a subset of observables, namely the mass, angular momentum and electric charge, as we know from the no-hair theorem. Therefore, this entropy satisfies our definition of a coarse-grained entropy (1.6). Of course, this is what we would expect from the interpretation of the Bekenstein-Hawking entropy as the thermodynamic entropy of the black hole.

## 5.2 The Central Dogma

The gravitational path integral is one of several routes to derive the Bekenstein-Hawking entropy. Strominger and Vafa [10] calculated the statistical mechanical entropy in string theory for a certain class of extremal Reissner-Nordström black holes. The result matches the Bekenstein-Hawking entropy exactly to leading order. The same result can be achieved in AdS/CFT: Strominger [50] considered black holes whose near-horizon geometries are locally three-dimensional anti-de Sitter space and calculated their statistical mechanical entropy on the CFT side of the duality. The result matches the Bekenstein-Hawking entropy exactly.

These results support the idea that the Bekenstein-Hawking entropy arises from hidden microstates of the black hole. This has been such an important idea in studies of the black hole information paradox that it has been named the ‘central dogma’ [21]. It can be summarized as follows:

As seen from the outside, a black hole can be described in terms of a quantum system with  $e^{A/4}$  microstates, which evolves unitarily under time evolution.

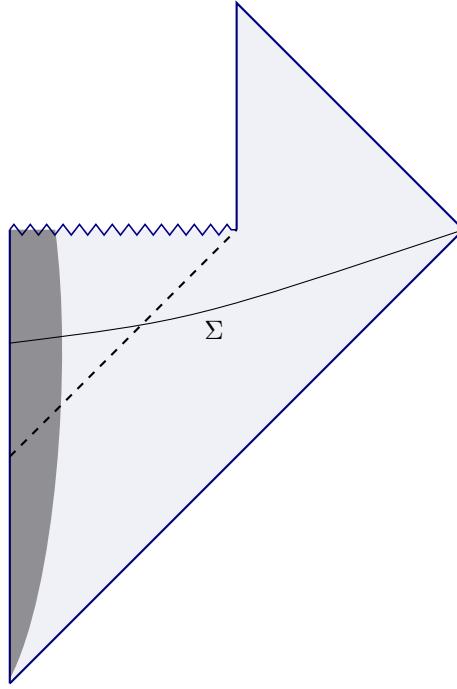


Figure 5.1: Partial Cauchy slice.

### 5.3 The Information Paradox

Let us now return to the Penrose diagram we had for the evaporating spacetime and the partial Cauchy slices we drew in figure 2.2. We draw a new partial Cauchy slice  $\Sigma$  which extends from  $r = 0$  to infinity at a time after the black hole has formed, as shown in figure 5.1. This Cauchy slice has the same domain of dependence as  $\Sigma_1$  in figure 2.2. We can therefore expect unitary evolution from  $\Sigma_1$  to  $\Sigma$ : If the black hole formed from an initially pure state on  $\Sigma_1$  we also have a pure state on  $\Sigma$ . Let us now divide  $\Sigma$  at the event horizon, such that  $\Sigma_{\text{bh}}$  is the part inside and  $\Sigma_{\text{rad}}$  is the part outside the horizon. The outside observer only has access to  $\Sigma_{\text{rad}}$ . The state available to such an observer is therefore described by tracing out  $\Sigma_{\text{bh}}$ , leaving a mixed state because of the entanglement between the matter on  $\Sigma_{\text{bh}}$  and  $\Sigma_{\text{rad}}$ .

If we believe in the central dogma, we can understand the matter on  $\Sigma_{\text{bh}}$  as a quantum system with thermodynamic (coarse-grained) entropy given by the Bekenstein-Hawking entropy  $S_{\text{BH}}$ . Let us denote the fine-grained entropy on  $\Sigma_{\text{bh}}$  as  $S_{\text{bh}}$  and the fine-grained entropy on  $\Sigma_{\text{rad}}$  as  $S_{\text{rad}}$ . Since the state on the full slice  $\Sigma$  is pure, we must have  $S_{\text{rad}} = S_{\text{bh}}$ . Recall that the fine-grained entropy of a system is bounded by its coarse-grained entropy (1.9). We must therefore have  $S_{\text{bh}} \leq S_{\text{BH}}$ .<sup>1</sup> It follows that

$$S_{\text{rad}} \leq S_{\text{BH}}. \quad (5.14)$$

The Bekenstein-Hawking entropy will decrease as the black hole shrinks during evaporation.  $S_{\text{rad}}$ , on the other hand, is the entropy of Hawking radiation. We know that this is described by a thermal density matrix at all times. The fine-grained entropy of Hawking radiation will therefore increase as more radiation is emitted. Therefore, we

<sup>1</sup>Note the difference between  $S_{\text{bh}}$  and  $S_{\text{BH}}$ . The lowercase bh stands for black hole, while the uppercase BH stands for Bekenstein-Hawking.

expect the inequality (5.14) to be broken at some time that we will call  $t_{\text{Page}}$ . *This* is the information paradox.

### 5.3.1 The Page Curve

We will now check when the inequality breaks. We have already found the time evolution of the black hole's mass  $M(t)$  in (2.21). Since the surface area of the Schwarzschild black hole is  $A = 4\pi r_s^2 = 16\pi M^2$  we can insert  $M(t)$  in (1.23) to get the time dependence of the Bekenstein-Hawking entropy:

$$S_{\text{BH}}(t) = 4\pi M_0^2 \left(1 - \frac{t}{t_{\text{evap}}}\right)^{2/3}. \quad (5.15)$$

We know that the density matrix of Hawking radiation is given by (2.16). We can insert this directly into the formula for the von Neumann entropy (1.3). Let us go ahead and do this following Page [51]. We write down (2.16) again for convenience:

$$\rho_{mn} = \delta_{mn} \frac{\Gamma^n (e^x \mp 1)^{\pm 1}}{(e^x \mp 1 \pm \Gamma)^{n \pm 1}}, \quad (5.16)$$

where the upper sign is for bosons, which can have any nonnegative integer integral number of quanta, and the lower sign is for fermions, which can only have  $n = 0$  or  $n = 1$ .  $\Gamma$  is the transmission coefficient of the hole to the particle of species  $s$  and rest mass  $\mu$  in a mode of energy  $x \equiv T\omega$ , helicity  $p$ , and angular momentum  $(l, m)$ . Let us now insert in (1.3) to find the von Neumann entropy  $S_{\omega,s,p,l,m} = \text{Tr}(-\rho_{\omega,s,p,l,m} \ln \rho_{\omega,s,p,l,m})$  of a single mode emitted at time  $t$ .

Assuming the black hole to be large, its temperature will be low and the emitted particles will be massless, consisting only of photons and gravitons. Thus we can take the boson equation of (2.16), which will be convenient to write as

$$\rho_{mn,\omega,s,p,l,m}(t) = \delta_{mn} z y^n \quad (5.17)$$

where  $z = \frac{e^x - 1}{e^x - 1 + \Gamma}$  and  $y = \frac{\Gamma}{e^x - 1 + \Gamma}$ . Note that  $z = y - 1$ . We also have  $y < 1$  because  $\omega > 0$ . Then we have

$$\begin{aligned} S_{\omega,s,p,l,m}(t) &= - \sum_{n=0}^{\infty} z y^n \ln(z y^n) \\ &= -z \sum_{n=0}^{\infty} y^n (\ln z + n \ln y) \\ &= -z \ln z \sum_{n=0}^{\infty} y^n - z \ln y \underbrace{\sum_{n=0}^{\infty} n y^n}_{y \frac{d}{dy} \sum_{n=0}^{\infty} y^n} \\ &= -z \ln z \frac{1}{1-y} - z y \ln y \frac{1}{(1-y)^2} \\ &= -\ln z - \frac{y \ln y}{z}, \end{aligned} \quad (5.18)$$

where we have used the sum of a geometric series  $\sum_{n=0}^{\infty} y^n = \frac{1}{1-y}$  when  $y < 1$ . Inserting for  $z$  and  $y$  again we arrive at

$$S_{\omega,s,p,l,m}(t) = \ln \left(1 + \frac{\Gamma_{\omega,s,p,l,m}}{e^x - 1}\right) + \frac{\Gamma_{\omega,s,p,l,m}}{e^x - 1} \ln \left(\frac{e^x - 1}{\Gamma_{\omega,s,p,l,m}} + 1\right). \quad (5.19)$$

Integrating over all modes and summing over particle types, we have that the total entropy of Hawking radiation  $S$  emitted at time  $t$  is

$$S(t) = \frac{1}{2\pi} \sum_{s,p,l,m} \int_0^\infty d\omega \left[ \ln \left( 1 + \frac{\Gamma_{\omega,s,p,l,m}}{e^x - 1} \right) + \frac{\Gamma_{\omega,s,p,l,m}}{e^x - 1} \ln \left( \frac{e^x - 1}{\Gamma_{\omega,s,p,l,m}} + 1 \right) \right]. \quad (5.20)$$

Since the emitted modes are perfectly thermal, there can be no correlation between modes emitted at different times. We can therefore set  $S(t)$  as the rate of change of the total entropy  $S_{\text{rad}}$  of all the radiation that has been emitted at time  $t$ :

$$\frac{dS_{\text{rad}}}{dt} = S(t). \quad (5.21)$$

The integral for  $S(t)$  was solved numerically by Page [37] for the case of photon and graviton emission, who found

$$\frac{dS_{\text{rad}}}{dt} = \frac{\gamma}{M(t)}, \quad (5.22)$$

where  $\gamma \approx 1.4 \cdot 10^{-3}$ . Inserting  $M(t)$  from (2.21), we get the differential equation

$$\frac{dS_{\text{rad}}}{dt} = \frac{\gamma}{M_0} \left( 1 - \frac{t}{t_{\text{evap}}} \right)^{-1/3}. \quad (5.23)$$

Solving with the initial condition that there is no radiation entropy when the black hole forms at  $t = 0$ , we obtain

$$S_{\text{rad}}(t) = \frac{\gamma M_0^2}{2\alpha} \left[ 1 - \left( 1 - \frac{t}{t_{\text{evap}}} \right)^{2/3} \right]. \quad (5.24)$$

Note that this is a fine-grained entropy obtained from a semiclassical approximation. Since the semiclassical approximation is accurate until a Planck time away from complete evaporation, we expect (5.24) to agree with the actual fine-grained entropy of the radiation until this time.

Now that we have  $S_{\text{rad}}(t)$  and  $S_{\text{BH}}(t)$  it is interesting to check if black hole evaporation satisfies the second law of thermodynamics. Taking the time derivative of (5.15) yields

$$\frac{dS_{\text{BH}}}{dt} = -\frac{8\pi\alpha}{M(t)}. \quad (5.25)$$

Inserting  $S_{\text{rad}}$  for  $S_{\text{matter}}$  into the generalized entropy (1.22) and using (5.25) and (5.22), we have

$$\frac{dS_{\text{gen}}}{dt} = \frac{dS_{\text{BH}}}{dt} + \frac{dS_{\text{rad}}}{dt} = \frac{\gamma - 8\pi\alpha}{M} \approx \frac{7.0 \cdot 10^{-5}}{M} > 0. \quad (5.26)$$

Thus the second law is satisfied.

Solving  $S_{\text{BH}}(t) = S_{\text{rad}}(t)$ , we find that inequality (5.14) breaks at the time  $t_{\text{Page}} \approx 0.54t_{\text{evap}}$ . A plot of  $S_{\text{BH}}(t)$  and  $S_{\text{rad}}(t)$  is shown in figure 5.2. Note that the black hole is still large at the Page time. The inequality is broken long before the black hole approaches complete evaporation and the assumptions needed to do QFT in curved spacetime would have broken down.

The Page time is named after Don Page, who studied how  $S_{\text{rad}}$  *should* evolve if the central dogma is true [37, 11, 12]. He found that the entropy will increase according to  $S_{\text{rad}}(t)$  up to the Page time  $t_{\text{Page}}$ , after which it will decrease, lying close to the maximum

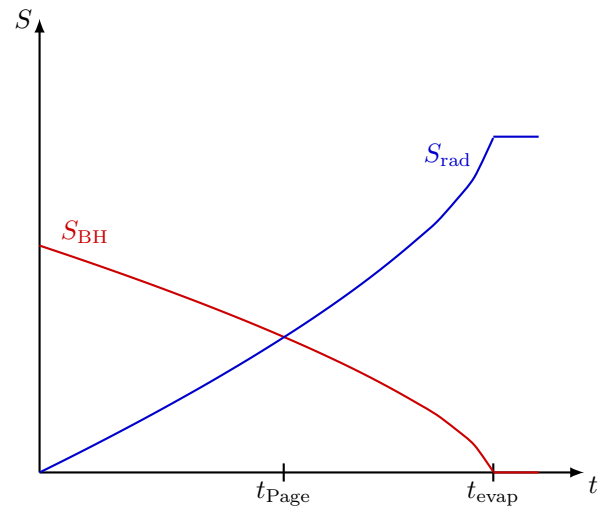


Figure 5.2: Time evolution of the Bekenstein-Hawking entropy (5.15) and the entropy of Hawking radiation (5.24).

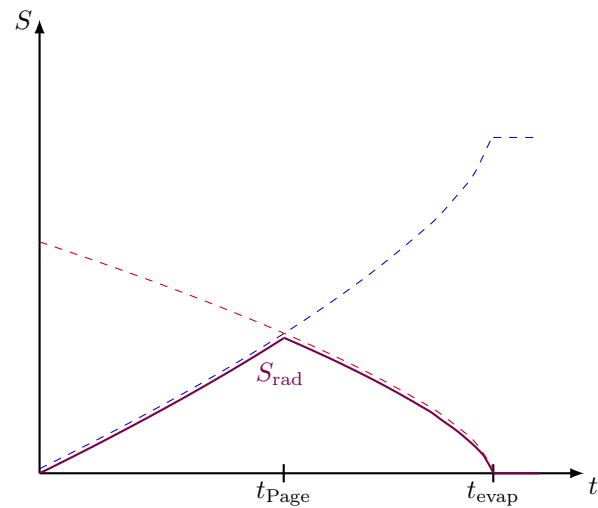


Figure 5.3: The Page curve for the entropy of Hawking radiation. Note that since the total state of the black hole and Hawking radiation is pure, the fine-grained entropy of the black hole is equal to that of the Hawking radiation. Thus the Page curve also applies to the black hole.

allowed entropy  $S_{\text{BH}}$ . The resulting curve is called the *Page curve*, and is shown in figure 5.3.

The Page curve can be understood physically as follows. At early times, the number of Hawking quanta will be low compared to the black hole quanta. The entanglement between the Hawking quanta and the black hole will therefore make the Hawking quanta appear to be in a maximally mixed, thermal state. This is true until the Page time. At this time the number of emitted Hawking quanta is roughly the same as the number of black hole degrees of freedom, and new emitted quanta are likely to be entangled with quanta that were emitted at earlier times. This causes the entropy to decrease. The entropy will keep decreasing until it reaches zero when the black hole has completely evaporated.



### 5.3.2 The Strength of the Information Paradox

To make it very clear, we summarize the black hole information paradox below:

The black hole information paradox is the conflict between

1. The central dogma, which implies that (5.14) must hold, meaning that the fine-grained entropy of Hawking radiation is bounded by the Bekenstein-Hawking entropy.
2. The result from QFT in curved spacetime that Hawking radiation is perfectly thermal and described by the density matrix (2.16), which implies that the fine-grained entropy of Hawking radiation is monotonically increasing.

Statement 1. and 2. are in conflict at times  $t \geq t_{\text{Page}} \approx 0.54t_{\text{evap}}$  because 2. implies that inequality (5.14) is broken at  $t = t_{\text{Page}}$ .

According to one definition [52], a paradox is an apparently impeccable argument to an impossible conclusion—such as a pair of apparently impeccable arguments whose conclusions contradict each other. By this definition, the black hole information paradox is a true paradox. We have every right to believe in the central dogma: We have precise derivations of the Bekenstein-Hawking entropy in low-energy quantum gravity, string theory and the AdS/CFT duality. On the other hand, QFT in curved spacetime provides equally compelling reasons to believe that black holes emit exactly thermal radiation. These two conclusions contradict each other when inequality (5.14) is broken, which happens at a time we expect the models with which we computed  $S_{\text{rad}}$  to work perfectly well. This can only happen if the evaporation process is non-unitary. Note that in contrast to the non-unitarity that was argued for by Hawking in his Breakdown of Predictability paper [3], this non-unitarity happens in the domain of dependence of the partial Cauchy slice  $\Sigma_1$  of figure 2.2, a regime in which we expect unitary time evolution.



## Chapter 6

# The Holographic Principle and the AdS/CFT Correspondence

In this chapter, we introduce the holographic principle and the AdS/CFT correspondence following Harlow [53]. We will be very brief and focus mostly on the conceptual aspects, as we will not perform any explicit calculations in AdS/CFT.

### 6.1 The Holographic Principle

Suppose we have a thermodynamic system with entropy  $S$  contained within a spherical region with surface area  $A$ . The total energy  $E$  of the thermodynamic system cannot exceed the mass  $M$  of a black hole of area  $A$ ; otherwise, it could not have been contained within the region. Now consider the process where we collapse a spherical shell of matter onto this system whose energy is  $M - E$ . This results in a black hole of mass  $M$  and surface area  $A$ , filling the spherical region. This black hole will have an entropy  $S_{\text{BH}} = A/4$ . For this process to satisfy the second law of thermodynamics, we must have  $S < S_{\text{BH}}$ . In other words, the maximum entropy of a region of space is proportional to the area of its boundary.

This is a peculiar property of gravity. One would naïvely guess that the maximum entropy of a region would be proportional to the *volume* of the region: One can always imagine putting more degrees of freedom into that volume. This would indeed have been the case if we had considered a system in quantum field theory without gravity [54]. Instead, gravity places a stricter boundary: If one tries the naïve procedure of filling the volume with more degrees of freedom, a black hole will form at some point, and its entropy will be determined by the surface area of its boundary. This led authors, most notably 't Hooft and Susskind [55, 56], to suggest that the physics taking place in a three-dimensional region can somehow be encoded on a two-dimensional surface surrounding that region. The idea is called the *holographic principle* and is believed to be a property of quantum gravity. It remained a vague idea until 1997, when Maldacena made a precise example with his AdS/CFT duality [13]. Here AdS stands for anti-de Sitter space and CFT stands for conformal field theory.

## 6.2 The AdS/CFT Correspondence

Anti-de Sitter space is the maximally symmetric solution of the field equations with a negative cosmological constant. In  $d + 1$  dimensions its metric can be written as [53]

$$ds^2 = - \left[ 1 + \left( \frac{r}{r_{\text{ads}}} \right)^2 \right] dt^2 + \frac{dr^2}{1 + \left( \frac{r}{r_{\text{ads}}} \right)^2} + r^2 d\Omega_{d-1}^2, \quad (6.1)$$

where  $t \in (-\infty, \infty)$ ,  $r \in [0, \infty)$  and  $r_{\text{ads}}$  is a characteristic length scale, related to the vacuum energy  $\rho_0$  as

$$\frac{1}{r_{\text{ads}}^2} = -\frac{16\pi G\rho_0}{d(d-1)}. \quad (6.2)$$

The spacetime has its own interesting boundary structure; it does not asymptotically approach Minkowski space. Introducing a new coordinate  $\tan \eta = r/r_{\text{ads}}$ , the metric is transformed to the form:

$$ds^2 = \frac{1}{\cos^2 \eta} \left[ -dt^2 + d\eta^2 + \sin^2 \eta d\Omega_{d-1}^2 \right], \quad (6.3)$$

where  $0 \leq \eta < \pi/2$ . We can conformally compactify by disregarding the diverging prefactor and including the conformal boundary at  $\eta = \pi/2$ . The resulting Penrose diagram is the cylinder shown in figure 6.1.

The Penrose diagram tells us that we should think of AdS as a box. We choose reflecting boundary conditions at  $r = \infty$ , which is necessary in order to think of AdS as a closed system. Then a massless particle sent out from the centre is able to propagate all the way to the boundary and back in a finite proper time as seen from an observer at the centre. A massive particle will not reach all the way to the boundary; it will travel a finite distance before returning to the centre in a finite proper time. These observations are formalized in the statement that the conformal boundary is timelike; it has the topology of  $\mathbb{R} \times \mathbb{S}^{d-1}$ . The induced metric at the conformal boundary is given by

$$ds^2 = -dt^2 + d\Omega_{d-1}^2. \quad (6.4)$$

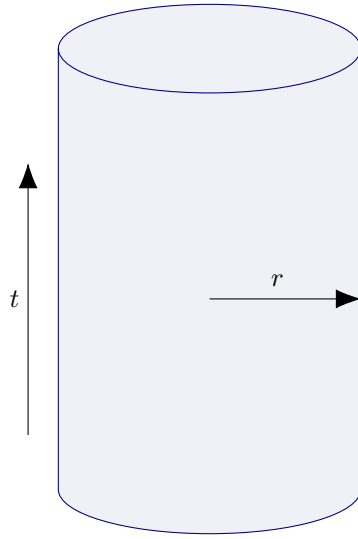
A conformal field theory is a quantum field theory that is invariant under conformal transformations; that is, transformations  $x^\mu \rightarrow x'^\mu(x)$  such that the new metric satisfies [57]

$$g_{\mu\nu}(x) \rightarrow \Omega^2(x)g_{\mu\nu}(x). \quad (6.5)$$

The set of spacetime transformations that satisfy this requirement is called the conformal group. The Poincaré group, which includes translations, boosts and rotations, is a subgroup of the conformal group. The conformal group also allows the special conformal transformation [53]

$$x^\mu \rightarrow \frac{x^\mu + a^\mu x^2}{1 + 2x_\nu a^\nu + a^2 x^2}. \quad (6.6)$$

In his original paper [13], Maldacena showed that a four-dimensional supersymmetric Yang-Mills theory is equivalent to type IIB string theory in an  $\text{AdS}_5$  background with the Yang-Mills theory on its boundary. The duality has been generalized since then. The present understanding is that the duality is valid outside the realm of string theory. The modern statement is the following:



**Figure 6.1:** The Penrose diagram of  $AdS_{d+1}$  for  $d = 2$ . The ‘bulk’ is the AdS region inside the cylinder. The CFT lives on the boundary.

Any relativistic conformal field theory on  $\mathbb{R} \times \mathbb{S}^{d-1}$  with metric (6.4) can be interpreted as a theory of quantum gravity in asymptotically  $AdS_{d+1} \times \mathcal{M}$  spacetime. Here  $\mathcal{M}$  is some compact manifold that may or may not be trivial.

Where by  $AdS_{d+1} \times \mathcal{M}$  we mean a spacetime  $\mathcal{M}$  that is asymptotically AdS. In loose terms one can say that the AdS/CFT duality tells us that an asymptotically AdS universe can be described by a CFT living on the boundary of that universe, thus making an explicit example of the holographic principle. AdS/CFT can be seen as a ‘dictionary’ that relates quantities on each side of the duality. In this way, one can use CFT on the boundary to perform certain calculations in the AdS region or vice versa. The AdS region is often referred to as the ‘bulk’. In figure 6.1 the bulk is the region inside the cylinder. The CFT lives on the conformal boundary of that cylinder.

### 6.3 AdS/CFT and the Information Paradox

The AdS/CFT duality is a conjecture that has not been proved. However, there is an enormous amount of evidence in its favour in the form of calculations matching on either side of the duality [58]. Given the validity of the conjecture, it tells us that black hole evaporation should somehow be unitary in the sense that it follows the Page curve, at least for a black hole in asymptotically AdS: Such a black hole can be equivalently described by a CFT on the AdS boundary, which is a manifestly unitary theory.

For this reason, the AdS/CFT duality shifted the general opinion to the belief that black hole evaporation must be unitary, and that the central dogma is true. The conjecture even convinced Hawking himself [14], who had been defending the opposite belief ever since the publication of his ‘Breakdown of Predictability in Gravitational Collapse’ [3].

If the central dogma is true, the monotonically increasing entropy of Hawking radiation that we found cannot be the final answer. The state for which we calculated the entropy is a state in the Hilbert space of QFT in curved spacetime,  $\mathcal{H}_{\text{QFT}}$ . This is

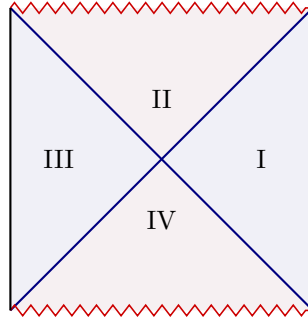


Figure 6.2: The Penrose diagram of the eternal black hole in asymptotically AdS.

not the Hilbert space of the full theory of quantum gravity  $\mathcal{H}_{\text{QG}}$ . We may therefore hope that the monotonically increasing entropy of Hawking radiation is only a fine-grained entropy of the state in  $\mathcal{H}_{\text{QFT}}$  and not in  $\mathcal{H}_{\text{QG}}$ . The AdS/CFT duality has opened the door to studying the entropy in quantum gravity, at least in asymptotically AdS spacetime, which we will do in the next chapter.

## 6.4 ER = EPR

Now that we have defined the AdS/CFT correspondence, we can continue the discussion of the thermofield double we started at the end of chapter 4. We noted that the Hartle-Hawking vacuum of the eternal black hole is equivalent to a thermofield double state (4.28), where the states  $|n\rangle_A$  and  $|n\rangle_B$  that enter are the states of observers at fixed coordinates outside the black hole. Let us now consider the eternal black hole in an asymptotically AdS spacetime. Its Penrose diagram is shown in figure 6.2. The spacetime has two boundaries: One at spatial infinity in region I, the other at spatial infinity in region III. On each boundary, we have a CFT. Maldacena [59] proposed that the eternal black hole in asymptotically AdS can also be described as two CFTs that are entangled in the thermofield double state (4.28).

Let  $O_I$  and  $O_{III}$  be operators in the CFT dual to region I and III, respectively. In the CFT description, the correlator  $\langle \text{TFD} | O_I O_{III} | \text{TFD} \rangle$  is non-zero because of the entanglement between the CFTs. The operators in the CFT are dual to operators in the bulk. In the bulk description, the correlator is non-zero because region I and III are connected by an Einstein-Rosen bridge (wormhole). In this sense, wormholes on the AdS side of the correspondence can be seen as entanglement on the CFT side of the correspondence. Maldacena and Susskind have later made the radical suggestion that this principle holds true outside of AdS/CFT, and that in quantum gravity, wormholes are equivalent to entanglement [60].

The authors note a number of similarities between the two phenomena. The simplest to state is the fact that both seem like strange violations of locality at first sight, but in both cases, they do not provide mechanisms for superluminal signal propagation. In the case of entanglement, no local operation on one member of the entangled pair can influence the other before a classical signal can propagate between them. In the case of wormholes, no signal can propagate through the wormhole from one exterior to the other, as can be seen in the Penrose diagram of the eternal black hole. The principle was given the slogan ‘ER = EPR’, referring to the interesting coincidence that the phenomena were introduced in 1935 by the same authors: Wormholes were first discovered in the eternal black hole by Einstein and Rosen [61], while entanglement was discussed in the famous

paper on the claimed incompleteness of quantum mechanics by Einstein, Podolsky and Rosen [62].





## Chapter 7

# Holographic Entanglement Entropy and Replica Wormholes

The discovery of the AdS/CFT duality sparked a change in the general opinion on the black hole information paradox. The community went from asking *if* information can escape a black hole to asking *how* it can happen. More specifically, the community has been asking how the entropy of Hawking radiation can follow the Page curve. There has recently been progress in answering this question, and concepts from AdS/CFT have been important contributors. Among these concepts is *holographic entanglement entropy*, which we will introduce in this chapter. A recently conjectured formula for the holographic entanglement entropy that includes quantum gravity effects has been applied to Hawking radiation. The formula gives the result one would hope for—that the entropy of Hawking radiation follows the Page curve. We will provide qualitative arguments as to how this happens. Finally, we will show how this result can be obtained from the gravitational replica trick without using the AdS/CFT duality. This is achieved by allowing for more exotic saddles in the gravitational path integral referred to as *replica wormholes*. The calculations involving replica wormholes are quite technical, and we will not attempt to do them here. We will instead give a more general argument which will be motivated by a simpler example.

### 7.1 Holographic Entanglement Entropy

Imagine that we are on the CFT side of the AdS/CFT duality; we have a CFT that lives on  $\mathbb{R} \times \mathbb{S}^{d-1}$ . In the Schrödinger picture, we can pick out a particular Cauchy slice and study the quantum state of the CFT on that slice. Consider the case where we decompose the slice into a region  $A$  and its complement  $\bar{A}$  and calculate the von Neumann entropy  $S_A$  of region  $A$  using the standard formula (1.3). How can we calculate that same entropy on the AdS side of the duality? This is answered by the Ryu-Takayanagi (RT) proposal for stationary spacetimes, later generalized by Hubeny, Rangamani and Takayanagi (HRT) to include non-stationary spacetimes [63, 64]. The RT/HRT proposal says that we can calculate  $S_A$  using the bulk gravity theory by the formula

$$S_A = \frac{\text{Area}(X)}{4}, \tag{7.1}$$

where  $X$  is the codimension-two extremal-area surface in the bulk geometry with the property that  $\partial X = \partial A$ . If there is more than one such  $X$  we choose the one with the

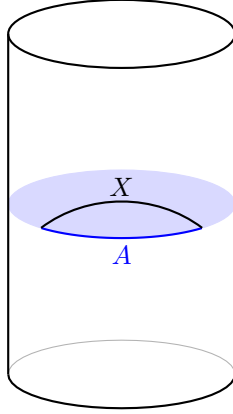


Figure 7.1: The RT/HRT proposal. The entanglement entropy of the reduced density matrix of region  $A$  in the CFT can be calculated by the area of the codimension-two extremal-area surface  $X$ . Note that since this is the Penrose diagram of  $AdS_3$  the codimension-two surface  $X$  is a line, meaning that we should insert the length for  $\text{Area}(X)$  in (7.1). We choose the surface with the shortest spatial length, or more generally the shortest spacelike geodesic if the spacetime is non-stationary.

smallest area. We also have the requirement that  $X$  is homologous to  $A$ , which means that we can continuously deform the curve to coincide with  $A$ .

There is an obvious similarity between this formula and the Bekenstein-Hawking entropy (1.23); the only difference is the surface we calculate the area of. This is no coincidence. In fact, we can view it as a generalization of the Bekenstein-Hawking entropy. We will later see how this formula can be motivated by a calculation similar to the one we did when we derived the Bekenstein-Hawking entropy from the gravitational path integral in chapter 5.

### 7.1.1 Entanglement Wedge Reconstruction

We may also ask where on the boundary a given bulk region is described. This can be answered by *entanglement wedge reconstruction*. The entanglement wedge of a region  $A$  can be defined in the following way [65]:

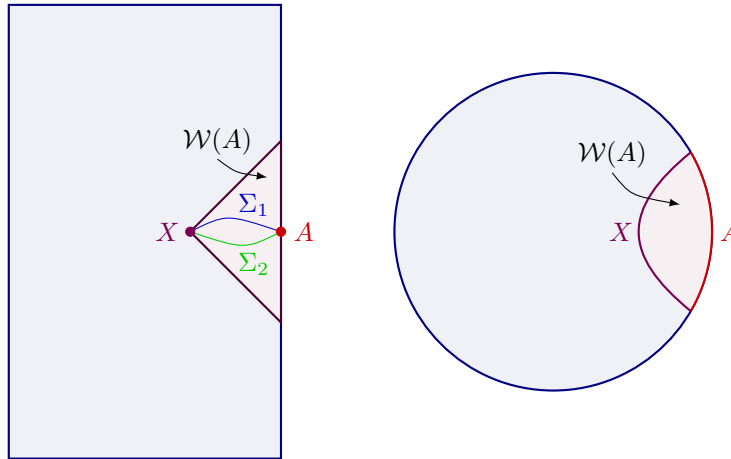
Find an extremal surface  $X$  in the RT/HRT proposal (7.1), then identify the subregion  $\Sigma$  of a bulk Cauchy slice, which is bound by  $X$  and  $A$ :  $\partial\Sigma = X \cup A$ . The entanglement wedge of  $A$ , denoted by  $\mathcal{W}(A)$ , is the bulk domain of dependence of  $\Sigma$ :

$$\mathcal{W}(A) = D(\Sigma). \quad (7.2)$$

A bulk operator can be reconstructed using a given boundary region  $A$  if, and only if, the bulk operator is contained in the entanglement wedge of  $A$  [16]. Put simpler, the physics in the entanglement wedge of  $A$  is encoded on  $A$ . The entanglement wedge is illustrated in figure 7.2.

### 7.1.2 Quantum Corrections

The RT/HRT formula is only valid to leading order  $\mathcal{O}(\hbar^{-1})$ . Engelhardt and Wall [15] have later generalized it to arbitrary orders in  $\hbar$ , allowing for quantum corrections. They



**Figure 7.2:** The entanglement wedge  $\mathcal{W}(A)$  of a region  $A$  on the boundary. The left figure shows the AdS cylinder as seen from the side. The right shows a time slice as seen from above. Since  $\Sigma_1$  and  $\Sigma_2$  are bounded by  $X$  and  $A$  and contained in  $\mathcal{W}(A)$ , they give the same entropy.

conjectured that the fine-grained entropy of the state in quantum gravity is given by

$$S_A = \min_X \left\{ \text{ext}_X \left[ \frac{\text{Area}(X)}{4} + S_{\text{semi-cl}}(\Sigma_X) \right] \right\}, \quad (7.3)$$

where  $X$  is a codimension-two surface,  $\Sigma_X$  is the region bounded by  $X$  and  $A$ , and  $S_{\text{semi-cl}}(\Sigma_X)$  is the von Neumann entropy of the quantum fields  $\Sigma_X$  appearing in the QFT in curved spacetime description. The surface  $X$  is the generalized Ryu-Takayanagi surface and is referred to as the ‘quantum extremal surface’ (QES). It is chosen such that it extremizes the generalized entropy. If there is more than one such surface, the one that minimizes the entropy is chosen. Note that for given  $X$ , any spacelike surface  $\Sigma$  in the entanglement wedge of  $A$  with boundary  $\partial\Sigma = X \cup A$  results in the same entropy, as shown in figure 7.2. Thus we can think of  $S_{\text{semi-cl}}(\Sigma_X)$  as the semiclassical entropy of the Hawking quanta in the entanglement wedge of  $\Sigma_X$ . We also note the striking resemblance to the generalized entropy (1.22); it is essentially the same formula for a more general surface.

The QES formula can be applied to black holes to find the time evolution of their entropy. We will later see that applying this formula to a black hole results in the Page curve, as we would expect for the unitary process of an initially pure state.

However, the Page curve we have considered is for the fine-grained entropy of Hawking radiation, not that of the black hole. One could of course argue that the fine-grained entropy of Hawking radiation must be equal to that of the black hole if we consider the system to be in an initially pure state. Nevertheless, we would be more convinced if we could perform a direct calculation of the fine-grained entropy of Hawking radiation and see that it does indeed follow the Page curve.

Luckily, there is a further generalization to the QES formula that allows us to do exactly that. It results from allowing the extremal surface to be disconnected, consisting of a radiation region  $\Sigma_{\text{Rad}}$  and a region inside the black hole referred to as the ‘island’,  $\Sigma_{\text{Island}}$  [18]:

$$S = \min_X \left\{ \text{ext}_X \left[ \frac{\text{Area}(X)}{4} + S_{\text{semi-cl}}(\Sigma_{\text{Rad}} \cup \Sigma_{\text{Island}}) \right] \right\}. \quad (7.4)$$

We will see that this formula gives a fine-grained entropy of Hawking radiation that follows a Page-curve exactly equal to the one we get for a black hole using (7.3).

If we want to use this formula to calculate the entropy of Hawking radiation, it seems that we must consider black holes in asymptotically Anti-de Sitter space rather than the asymptotically flat spacetimes we have considered so far. This was done by Penington [16], who found that the entropy does indeed follow the Page curve. However, the present understanding of the island formula is that it is much more general than the Ryu-Takayanagi formula; it requires neither the AdS/CFT duality nor even AdS spacetime. In fact, we will see that the entropy given by the island formula can be derived from the gravitational replica trick. Therefore we interpret it as a general formula for the fine-grained entropy of quantum systems coupled to gravity.

This means we can follow the schematic calculation of [21] and use the formula for the more familiar spacetime of a black hole formed from collapse in asymptotically flat spacetime. We will add a codimension-one cutoff surface outside the black hole that is going to play the role of the AdS boundary, and make sure that it is placed a couple of Schwarzschild radii outside the black hole such that its effective potential is inside the surface. The entropy we calculate will depend on the time of an observer at the cutoff surface.

First, we apply the QES formula (7.3) to find the entropy of the black hole. Then we apply the island formula (7.4) to find the entropy of Hawking radiation.

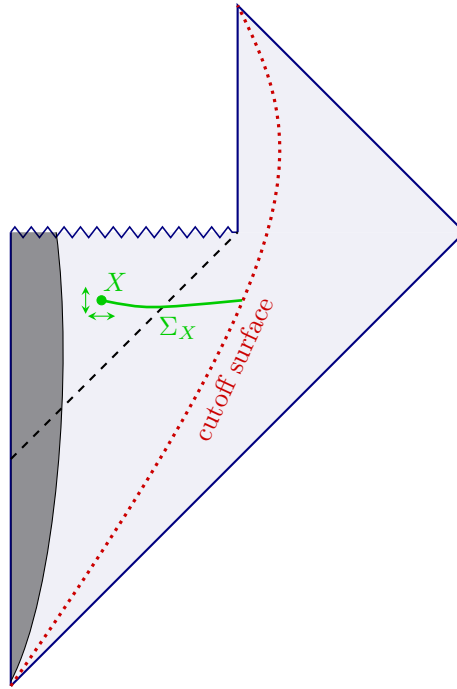
## 7.2 Applying Quantum Extremal Surfaces to Black Holes

Figure 7.3 shows the Penrose diagram of the spacetime and its cutoff surface. The surface  $X$  starts at the cutoff surface and is deformed into the enclosed region until an extremum is found. We refer to the region inside the cutoff as the ‘black hole region’ and the region outside as the ‘radiation region’.

The simplest surface that extremizes the generalized entropy is the empty surface. Choosing this surface gives an entropy  $S_{\text{semi-cl}}(\Sigma_X)$ , where  $\Sigma_X$  is the surface bounded by the origin and the cutoff surface. When the black hole forms and starts radiating, the entropy will stay zero because the outgoing Hawking quanta and its interior partner are both encapsulated by the cutoff surface. Once the outgoing quanta have escaped the cutoff surface, it is no longer contained in the entanglement wedge, while its interior partner is. This causes the entropy to increase. The semiclassical entropy on  $\Sigma_X$  must then be equal to the semiclassical entropy of the fields outside the cutoff surface. We have already found its time dependence: It is given by (5.24). The empty surface is extremal because if we had moved it slightly to be non-empty, the area term would increase, and so would the generalized entropy. Thus the black hole entropy will rise according to (5.24) in early stages of the evaporation.

This will not go on indefinitely: There is a second extremal surface that lies close to the event horizon. Its position at time  $t$  is found by going back along the cutoff surface by a time of order  $r_s \log S_{\text{BH}}$  and shooting off an ingoing light ray. Then the surface lies close to the point where the light ray intersects the horizon. Thus we get contributions both from the area term and  $S_{\text{semi-cl}}(\Sigma_X)$ . The area term will dominate as there are few Hawking quanta in the entanglement wedge of  $\Sigma_X$ . Therefore the contribution from this extremal surface will lie close to the Bekenstein-Hawking entropy, and the time evolution will be decreasing as we found in (5.15).

In order to prove that this is indeed an extremal surface, one must show that the change of area of  $X$  under a small deformation in any direction perfectly balances the



**Figure 7.3:** Penrose diagram of the spacetime under consideration and its cutoff surface. In order to find the entropy of the black hole at time  $t$  we place  $X$  at the cutoff surface at time  $t$  and deform it into the enclosed region until an extremum is found. Here  $\Sigma_X$  is the region bounded by  $X$  and the cutoff surface. Note that the cutoff surface is codimension-one.

change in  $S_{\text{semi-cl}}(\Sigma_X)$ . To get some intuition, consider the extremizing along the ingoing null direction. The area will be monotonically decreasing in this direction. The same is not true for  $S_{\text{semi-cl}}$ . We can see this by starting with  $X$  at the horizon and moving it inwards. As we move the surface inwards, we will include ingoing Hawking quanta that are entangled with outgoing quanta between the event horizon and the cutoff surface. This causes  $S_{\text{semi-cl}}$  to decrease. However, at some point, we will instead start to include ingoing quanta that are entangled with outgoing quanta outside the cutoff surface. This will increase  $S_{\text{semi-cl}}$ . In this regime, we can find a point where the changes in area and semiclassical entropy cancel each other out.

Finally, we take the minimum of the two extremal surfaces. This gives an entropy exactly equal to the Page curve, as shown in figure 7.4.

### 7.3 Applying the Island Formula to Hawking Radiation

Now we apply (7.4) to find the entropy of Hawking radiation. Here  $\Sigma_{\text{Rad}}$  is the region from the cutoff surface to infinity and  $\Sigma_{\text{Island}}$  is the region bounded by any quantum extremal surface and the origin. We can in principle have any number of islands, including zero. To find the island, we extremize the generalized entropy with respect to the position of the quantum extremal surface  $X$ . If there is more than one extremal surface, we choose the one that minimizes the entropy, just like we did for the black hole entropy. The setup is shown in figure 7.5.

Again, the simplest extremal surface is the empty surface, giving an entropy equal to the semiclassical entropy of Hawking radiation (5.24).

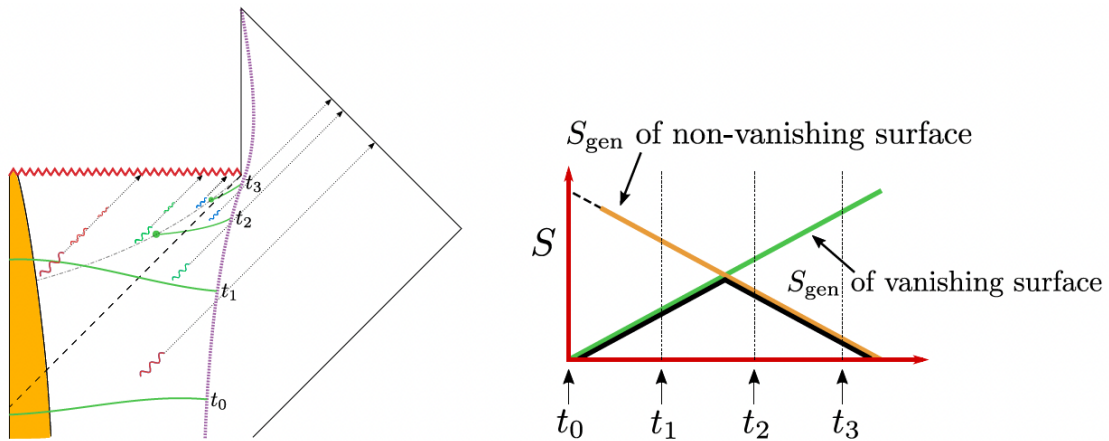


Figure 7.4: At early times, the quantum extremal surface that minimizes the generalized entropy is the empty surface. This causes the entropy to increase, following the green curve. At the Page time, the minimizing surface changes to the one just behind the horizon, causing the entropy to decrease along the yellow curve. The resulting generalized entropy thus follows the Page curve in figure 5.3 exactly. Adapted from [21].

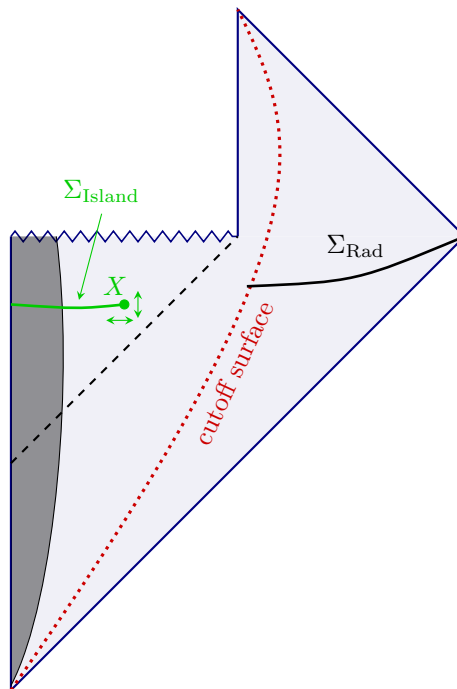
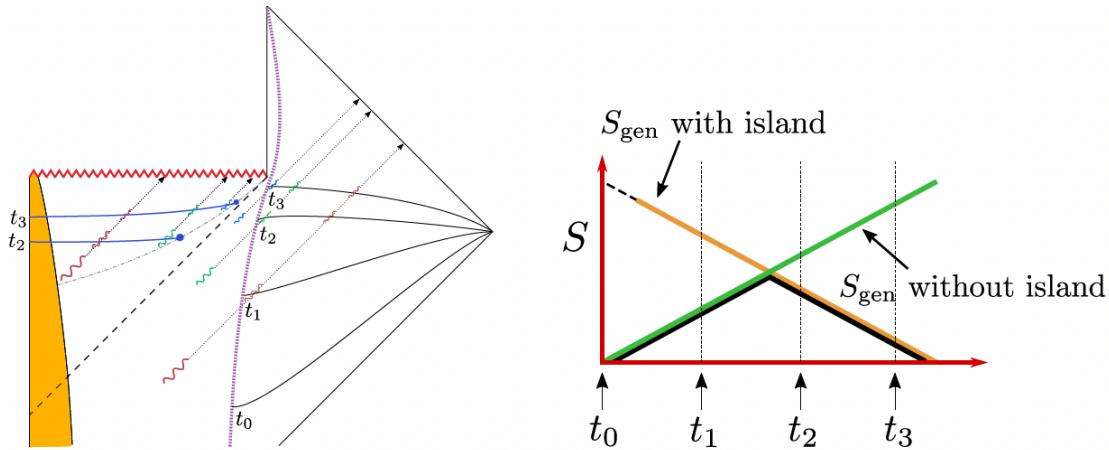


Figure 7.5: The setup for calculating the entropy of Hawking radiation. We evaluate the semiclassical entropy in the union of two disconnected regions  $\Sigma_{\text{Rad}}$  and  $\Sigma_{\text{Island}}$ . The quantum extremal surface  $X$  is again found by placing it at the cutoff surface and deforming it into the enclosed region until an extremum is found.



**Figure 7.6:** At early times, the quantum extremal surface that minimizes the generalized entropy is the empty surface. This causes the entropy to increase, following the green curve. At the Page time, an island contribution appears, causing the entropy to decrease along the yellow curve. The resulting generalized entropy thus follows the Page curve in figure 5.3 exactly. Adapted from [21].

We also have a non-empty extremal surface, giving an island contribution. This surface perfectly coincides with the non-empty surface we found when calculating the black hole entropy, lying just behind the horizon. To see this, consider the change in entropy as we move the surface slightly. If we move it further out, we will include modes that are entangled with modes inside the cutoff surface. Since these modes are not part of  $\Sigma_{\text{Rad}}$ , this will cause an increase in  $S_{\text{semi-cl}}$ . Similarly, if we move the surface inwards, we exclude modes that are entangled with modes in the entanglement wedge of  $\Sigma_{\text{Rad}}$ . This also causes an increase in  $S_{\text{semi-cl}}$ . Moving the surface inwards causes a decrease in the area contribution. Thus we can adjust it such that the changes in area and semiclassical entropy cancel each other out. Just like the extremal surface in the black hole calculation, this surface will be dominated by the area term. Thus the island contributes an entropy close to the Bekenstein-Hawking entropy (5.15).

Taking the minimum of the extremal surfaces, we are left with an entropy that follows the Page curve exactly as shown in figure 7.6.

## 7.4 Entanglement Wedges in Black Hole Evaporation

We can get a better understanding of the information escape from the black hole by studying the entanglement wedges of the black hole and the Hawking radiation at different times.

Before the Page time, the QES is the empty surface. Therefore, the entanglement wedges of the black hole and the Hawking radiation will cover the black hole region and the radiation region. This is shown in figure 7.7 (a). The entire black hole region is described by the black hole degrees of freedom. No information about the black hole region is contained in the radiation, just like we found in the semiclassical calculations.

This changes when the non-empty QES appears after the Page time. This surface was the same for the black hole calculation and the radiation calculation. We may therefore draw the entanglement wedges as in figure 7.7 (b). Now the black hole region is divided into two wedges; one that describes the black hole degrees of freedom, and the

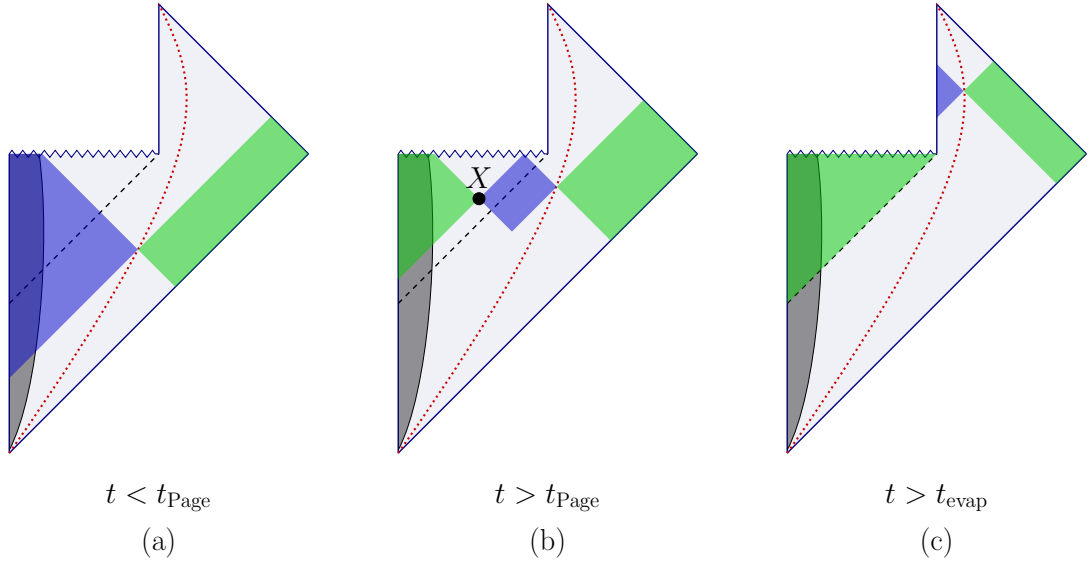


Figure 7.7: The entanglement wedges at different times.

island region that describes radiation degrees of freedom. We say that the island is part of the radiation and not of the black hole.

Finally, after the evaporation time, the entanglement wedges will be as in figure 7.7 (c), assuming that we can use the formulas up to the endpoint of evaporation. Now the region inside the cutoff surface is flat space, giving an empty QES. The entanglement wedge of the radiation includes the entire black hole interior, meaning that the entire black hole interior is part of the radiation.

The entanglement wedges do help us understand the process of information escape, but they do little to help us understand *how* it happens. To better understand this, we must see how we can use gravitational path integrals to derive the island formula (7.4).

## 7.5 The Replica Trick on the Cigar and a Proof of the RT Proposal

Before we show the calculations with replica wormholes, we warm up with a calculation inspired by the paper by Lewkowysz and Maldacena [66]. They used the replica trick to prove the RT proposal (7.1). This calculation can be seen as a generalization of the one we did when we calculated the black hole entropy in chapter 5. We will apply their method to that same system. That is, we will approximate the partition function by including the first term in (3.38):

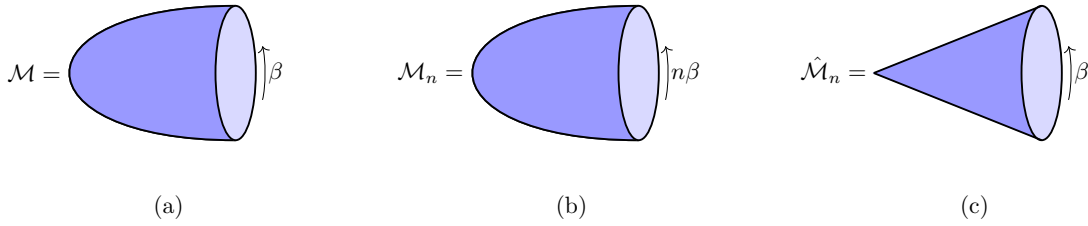
$$\log Z \approx -I_E[\mathcal{M}] = -I_G[\mathcal{M}], \quad (7.5)$$

where  $\mathcal{M}$  is the Euclidean Schwarzschild spacetime (the cigar) of figure 7.8 (a). Note that when we write  $I_G[\mathcal{M}]$  we really mean the gravitational action of the manifold  $\mathcal{M}$  with metric  $g$ , which we previously wrote as  $I_G[g]$ .

Following Lewkowysz and Maldacena, we will prepare a density matrix using the

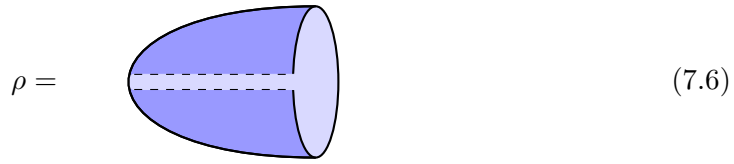


## 7.5. The Replica Trick on the Cigar and a Proof of the RT Proposal



**Figure 7.8:** The spacetime manifolds we construct. (a) is the cigar we use to construct the state in (7.6). (b) is the replicated manifold we get from glueing each copy of  $\rho$  defined in (7.6) cyclically together. (c) is the manifold we are left with after removing the replica symmetry  $\mathbb{Z}_n$  from (b).

cigar. We do so by making two cuts around  $\tau = 0$  from  $r = r_s$  to infinity:



We emphasize that there are no matter fields on this manifold. Any entropy of the density matrix will therefore be an entropy of the geometry itself. Just like before, we expect the result to be the Bekenstein-Hawking entropy (1.23).

We make a replicated manifold by glueing these cigars cyclically together. This results in a manifold  $\mathcal{M}_n$  that looks exactly like the cigar, but has period  $n\beta$ , as shown in figure 7.8 (b). Inserting in the replica trick formula (3.29), we get that the first-order contribution to the von Neumann entropy of the black hole is

$$S = \lim_{n \rightarrow 1} \partial_n (I_G[\mathcal{M}_n] - nI_G[\mathcal{M}]). \quad (7.7)$$

As discussed in section 3.3, we can now use the replica symmetry  $\mathbb{Z}_n$  to construct the quotient space  $\hat{\mathcal{M}}_n \equiv \mathcal{M}_n/\mathbb{Z}_n$ . This results in a geometry topologically equivalent to the original manifold  $\mathcal{M}$  where  $\tau$  now has period  $\beta$ . Since we consider a manifold with no matter fields on it, we do not have to worry about twist fields. However, the tip of the cigar at  $r = r_s$  is a fixed point under  $\mathbb{Z}_n$ . This results in a conical singularity at  $r = r_s$  on  $\hat{\mathcal{M}}_n$  with deficit angle

$$\Delta\phi = 2\pi \left(1 - \frac{1}{n}\right). \quad (7.8)$$

The manifold  $\hat{\mathcal{M}}_n$  is shown in figure 7.8 (c), where we have exaggerated the conical shape. We can generally take the fixed point to be a codimension-two surface. We can treat the surface as a source of energy-momentum which backreacts on  $\mathcal{M}$  to deform it to  $\hat{\mathcal{M}}_n$ . This is achieved by putting a *cosmic brane* at the fixed point [67]. The cosmic brane is a codimension-two surface with action given by

$$I_c = T_n \int_{\Sigma_{d-2}} d^{d-2}y \sqrt{h}, \quad (7.9)$$

where  $T_n = \frac{1}{4} \left(1 - \frac{1}{n}\right)$  is the tension of the cosmic brane and  $h$  is the determinant of the induced metric on the codimension-two surface. Thus we have that the metric on  $\hat{\mathcal{M}}_n$  must be a solution under variation of the sum of the gravitational and cosmic brane action on  $\mathcal{M}$ ,

$$I_G[\hat{\mathcal{M}}_n] = I_G[\mathcal{M}] + I_c[\mathcal{M}]. \quad (7.10)$$

Since the gravitational action is local, the gravitational action on  $\mathcal{M}_n$  should be  $n$  times that of a single domain:

$$I_G[\mathcal{M}_n] = nI_G[\hat{\mathcal{M}}_n]. \quad (7.11)$$

Note that the above procedure has turned the integer value  $n$  into a parameter in the cosmic tension, making it possible to analytically continue it to non-integer values. Thus we are now able to take the limit  $n \rightarrow 1$ . We can insert in (7.7) to get

$$\begin{aligned} S &= \lim_{n \rightarrow 1} \partial_n (nI_G[\hat{\mathcal{M}}_n] - nI_G[\mathcal{M}]) \\ &= \lim_{n \rightarrow 1} (I_G[\hat{\mathcal{M}}_n] + n\partial_n I_G[\hat{\mathcal{M}}_n] - I_G[\mathcal{M}]) \\ &= \lim_{n \rightarrow 1} (\partial_n I_G[\hat{\mathcal{M}}_n]), \end{aligned} \quad (7.12)$$

where we used that  $I_G[\hat{\mathcal{M}}_1] = I_G[\mathcal{M}_1] = I_G[\mathcal{M}]$  from (7.11).

Now, the key idea to evaluate this is to view  $\hat{\mathcal{M}}_n$  as a family of classical solutions to the equations of motion we get by varying the action in (7.10) and taking the derivative  $\partial_n$  as equivalent to varying the solution and its boundary conditions on the cosmic strings. Thus we only get contributions to  $S$  from the boundary term. We follow the suggestion by [67] and set up a physical boundary at  $r = r_0$  just outside the conical singularity. This means that the boundary term that contributes to  $S$  must be the GHY boundary term<sup>1</sup>:

$$\partial_n I_G[\hat{\mathcal{M}}_n] = -\partial_n I_B[\hat{\mathcal{M}}_n] = -\frac{1}{8\pi} \partial_n \oint_{r=r_0} \epsilon K \sqrt{h} d^3 y. \quad (7.13)$$

Since the boundary is set just outside the horizon, we can consider the metric of  $\hat{\mathcal{M}}_n$  in the near-horizon limit. We have already done this for the Euclidean Schwarzschild metric  $\mathcal{M}$  in chapter 4. We found that the near-horizon metric was given by (4.37), where  $\tau \in [0, 4\pi r_s)$ . It will be convenient to change to the angular coordinate  $\varphi = \frac{\tau}{2r_s}$ , such that the metric is transformed to

$$ds^2 = \rho^2 d\varphi^2 + d\rho^2 + r_s^2 d\Omega^2, \quad (7.14)$$

where  $\varphi \in [0, 2\pi)$ . We saw that the metric on  $\mathcal{M}_n$  has a Euclidean time period of  $n$  times that of  $\mathcal{M}$ . Therefore, let us define yet another Euclidean time coordinate given by  $\psi = n\varphi$ , transforming the metric into

$$ds^2 = \frac{\rho^2}{n^2} d\psi^2 + d\rho^2 + r_s^2 d\Omega^2, \quad (7.15)$$

where  $\psi \in [0, 2n\pi)$ . Finally, let us define a new radial coordinate  $r = \rho/n$  to transform the metric into

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu = r^2 d\psi^2 + n^2 dr^2 + r_s^2 d\Omega^2. \quad (7.16)$$

If we change the Euclidean time coordinate to have period  $2\pi$ , that is,  $\psi \in [0, 2\pi)$ , we are left with a metric with a conical singularity that describes  $\hat{\mathcal{M}}_n$ .

Let us now calculate the extrinsic curvature  $K$  on  $r = r_0$  of (7.16), for which  $\epsilon = 1$ . We have  $\sqrt{g} = n r r_s^2 \sin \theta$ . The induced metric on  $r = r_0$  is given by

$$ds^2 = h_{ab} dy^a dy^b = r_0^2 d\psi^2 + r_s^2 d\Omega^2, \quad (7.17)$$

<sup>1</sup>Note that this time we are taking the limit  $r_0 \rightarrow r_s$ , where the action of flat spacetime does not diverge. Therefore we do not need to include the  $K_0$  term to make the action finite.

with  $\sqrt{h} = r_0 r_s^2 \sin \theta$ . The normal vector calculated from (A.4) is  $n^\mu = \frac{1}{n}(0, 1, 0, 0)$ . Thus we get

$$K = \nabla_\mu n^\mu = \frac{1}{\sqrt{g}} \partial_\mu (\sqrt{g} n^\mu) \Big|_{r=r_0} = \frac{1}{r} \partial_r \left( \frac{r}{n} \right) \Big|_{r=r_0} = \frac{1}{nr_0}. \quad (7.18)$$

Now we can insert in (7.13) to get

$$\begin{aligned} \partial_n I_G[\hat{\mathcal{M}}_n] &= -\frac{1}{8\pi} \partial_n \int_0^{2\pi} d\psi \int_0^\pi d\theta \int_0^{2\pi} d\phi \frac{1}{nr_0} r_0 r_s^2 \sin \theta d^3 y \\ &= \frac{1}{n^2} \pi r_s^2 \\ &= \frac{1}{n^2} \frac{A}{4}. \end{aligned} \quad (7.19)$$

Inserting in (7.12), we are left with

$$S = \frac{A}{4} = S_{\text{BH}}, \quad (7.20)$$

as expected.

From here one can generalize the method, as Lewkowycz and Maldacena did, to a more general class of manifolds. In general, fixed points of the replica symmetry  $\mathbb{Z}_n$  appear as codimension-two surfaces and can be incorporated into the original manifold  $\mathcal{M}$  as cosmic branes that backreact and deforms it to  $\hat{\mathcal{M}}_n$ . The area formula for the entropy can then be viewed as arising from the action of the cosmic brane, with a minimal area condition from minimizing the action. In AdS/CFT situations, this entropy corresponds to the gravitational dual of the entropy of a subregion of the CFT on the boundary. From this, the extremal-area formula of the RT proposal (7.1) follows.<sup>2</sup>

## 7.6 Replica Wormholes

Now we will finally motivate how replica wormholes can be used to derive the island formula (7.4) for the entropy of Hawking radiation. In the previous calculation, we chose a saddle of the gravitational path integral and calculated its entropy. This time we will let gravity act dynamically; we will only specify the boundary condition and leave the rest for gravity to fill in. We will include the contribution from two saddles. One of them will give an entropy that corresponds to the case where there is no island, which we will refer to as the ‘Hawking saddle’. The other will correspond to the case where there forms an island. We will refer to this as the ‘replica wormhole saddle’ for reasons that will become apparent.

We will follow the general discussion of replica wormholes in the paper by Almheiri, Hartman, Maldacena, Shaghoulian and Tajdini [19] and show briefly how it applies to a model in which they perform explicit calculations. They consider a version of the information paradox where a black hole in two-dimensional AdS radiates into an attached Minkowski region. The Hawking radiation is modelled by a CFT that lives in both the gravitational and the non-gravitational region.<sup>3</sup> The setup is shown in figure

<sup>2</sup>We emphasize that the example we showed only serves to motivate this result. For a full derivation, see the paper by Lewkowycz and Maldacena [66].

<sup>3</sup>It might look like this calculation uses the AdS/CFT correspondence. This is not the case. The QFT is defined as a CFT only to simplify calculations.

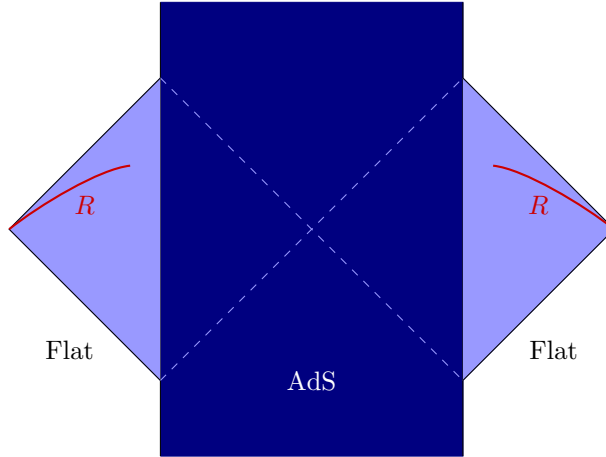


Figure 7.9: The setup. A black hole in asymptotically AdS with flat Minkowski spacetime attached to the AdS boundary.

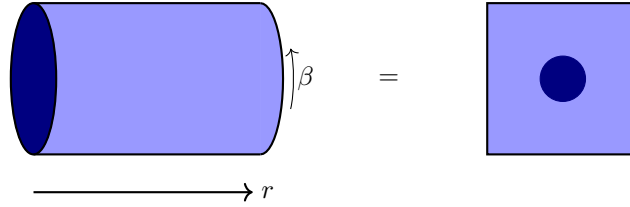


Figure 7.10: Boundary condition for the gravitational path integral. The light blue part is the non-gravitational region and the dark blue part is the gravitational region that will be filled in by the gravitational path integral.

7.9. Note that we use the eternal black hole, which does not evolve in time. The entropy we get will still have a time dependence because the region  $R$  on which we calculate the entropy evolves in time, as shown in the figure. We take the action to be  $I_E[g, \phi] = I_G[g] + I_{\text{CFT}}[g, \phi]$  where  $g$  denotes the metric and  $\phi$  denotes the matter fields. We will simplify our notation and simply write

$$I[\mathcal{M}] = I_G[\mathcal{M}] + I_{\text{CFT}}[\mathcal{M}], \quad (7.21)$$

where  $I$  is now the Euclidean action and we write all our actions as a functional of the manifold  $\mathcal{M}$  we consider,  $I[\mathcal{M}]$ . The CFT action  $I_{\text{CFT}}[\mathcal{M}]$  refers to the matter action on the manifold  $\mathcal{M}$ .

We prepare a boundary condition for the gravitational path integral as shown in figure 7.10. The light blue part is the non-gravitational region and the dark blue part is the gravitational region that will be filled in by the gravitational path integral. We represent the same path integral in the flat drawing on the right side of the equality. We do this because we want to make cuts on both sides of the eternal black hole; recall our discussion in section 4.5, where we noted that the spacelike hypersurface  $\tau = 0$  corresponds to region I of the eternal black hole, and halfway around, at  $\tau = \beta/2$ , corresponds to region III. We refer to the manifold prepared by this boundary condition as  $\mathcal{M}$ .

Now we make two disconnected cuts in the non-gravitational region to define the

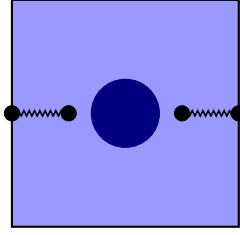


Figure 7.11: The boundary condition for  $\hat{\mathcal{M}}_n$ . The zigzagged lines show the cuts made in region  $R$  that are glued together. The black dots represent the insertion of twist operators.

reduced density matrix

$$\rho_R = \text{[Diagram of a square region with a central dark blue circle and two horizontal dashed lines extending from the left and right sides of the circle to the boundaries of the square.]} \quad (7.22)$$

Since the reduced density matrix lives in the non-gravitational region it corresponds to the density matrix of Hawking radiation. We will now use the replica trick to calculate this entropy while we include all possible saddles of the gravitational path integral that satisfy this boundary condition. As before, we construct the replica manifold  $\mathcal{M}_n$  by glueing each copy of  $\rho_R$  together cyclically along the cuts. The action on  $I[\mathcal{M}_n]$  is of course equivalent to the one for  $\mathcal{M}$  in (7.21),

$$I[\mathcal{M}_n] = I_G[\mathcal{M}_n] + I_{\text{CFT}}[\mathcal{M}_n]. \quad (7.23)$$

We will consider the contribution from each of the saddle points that satisfy the boundary condition of the gravitational path integral separately. As in the previous calculation, we will only consider the first term in the semiclassical approximation (3.37),

$$\log Z_n \approx -I[\mathcal{M}_n] = -I_G[\mathcal{M}_n] - I_{\text{CFT}}[\mathcal{M}_n], \quad (7.24)$$

and correspondingly for  $\mathcal{M}$ . Inserting in (3.29), we get

$$S_R = \lim_{n \rightarrow 1} \partial_n (I[\mathcal{M}_n] - nI[\mathcal{M}]). \quad (7.25)$$

The replicated manifold will of course have the replica symmetry  $\mathbb{Z}_n$ . Again, we construct the quotient space  $\hat{\mathcal{M}}_n = \mathcal{M}_n/\mathbb{Z}_n$ . This manifold is topologically equivalent to  $\mathcal{M}$  and has  $n$  copies of the CFT on it. As discussed in section 3.3 we must place twist operators at the boundary of the regions that are glued together. Figure 7.11 shows the boundary condition for  $\hat{\mathcal{M}}_n$ , where the black dots represent the insertion of twist operators. Whether we get a fixed point or not will now depend on how we fill in the gravitational region. If there is a fixed point, we must insert cosmic branes and twist operators at the conical singularities.

As in the previous calculation, we can use that the gravitational action is local to infer that the gravitational action of  $\mathcal{M}_n$  should be  $n$  times that of a single domain,

$$I_G[\mathcal{M}_n] = nI_G[\hat{\mathcal{M}}_n]. \quad (7.26)$$

Using (7.23), we get

$$I[\mathcal{M}_n] = nI_G[\hat{\mathcal{M}}_n] + I_{\text{CFT}}[\mathcal{M}_n]. \quad (7.27)$$

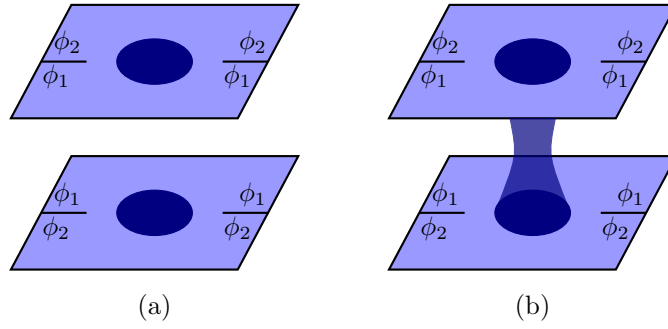


Figure 7.12: The two saddles  $\mathcal{M}_n$  for  $n = 2$ . The  $\phi_i$  are meant to show how the replicas are glued together in the sense of figure 3.2.

Now we can insert in (7.25) to get

$$S_R = \lim_{n \rightarrow 1} \partial_n \left( nI_G[\hat{\mathcal{M}}_n] + I_{\text{CFT}}[\mathcal{M}_n] - nI_G[\mathcal{M}] - nI_{\text{CFT}}[\mathcal{M}] \right). \quad (7.28)$$

The simplest saddle  $\mathcal{M}_n$  we can get is the one where we fill in the gravitational region on each replica separately. For  $n = 2$  this can be drawn as in figure 7.12 (a). Constructing  $\hat{\mathcal{M}}_n$  from this saddle does not give any fixed points. This means that no singularity forms in  $\hat{\mathcal{M}}_n$  and we can construct  $\hat{\mathcal{M}}_n$  directly from  $\mathcal{M}$ . Thus we can set

$$I_G[\hat{\mathcal{M}}_n] = I_G[\mathcal{M}]. \quad (7.29)$$

Inserting this in (7.28), we get

$$S_R^{\text{H}} = \lim_{n \rightarrow 1} \partial_n (I_{\text{CFT}}[\mathcal{M}_n] - nI_{\text{CFT}}[\mathcal{M}]). \quad (7.30)$$

But this is simply the replica trick expression for the von Neumann entropy of the CFT in the radiation region. Therefore we set

$$S_R^{\text{H}} = S_{\text{semi-cl}}(\rho_{\text{CFT}}(R)). \quad (7.31)$$

This is completely equivalent to the Hawking calculation of the radiation entropy, so nothing new happens here. This is the saddle we call the ‘Hawking saddle’, with the H in  $S_R^{\text{H}}$  standing for Hawking. Just like for the black hole in asymptotically flat spacetime,  $S_R^{\text{H}}$  is monotonically increasing.

The situation is different if we allow for more interesting topologies. That is, we can let the gravitational regions of the different replicas be connected as in figure 7.12 (b). This connection is what we refer to as the replica wormhole. In this case, there *are* fixed points under  $\mathbb{Z}_n$ . As in the previous calculation, we must therefore include cosmic branes at the fixed points in the gravitational action for  $\hat{\mathcal{M}}_n$ :

$$I_G[\hat{\mathcal{M}}_n] = I_G[\mathcal{M}] + I_c[\mathcal{M}], \quad (7.32)$$

in addition to twist operators at the same points. We can draw the manifold  $\mathcal{M}_n$  and  $\hat{\mathcal{M}}_n$  as in figure 7.13. Note however that in the trivial case  $n = 1$  no cosmic brane is needed and

$$I_G[\hat{\mathcal{M}}_1] = I_G[\mathcal{M}]. \quad (7.33)$$

Now we can insert in (7.28) to get

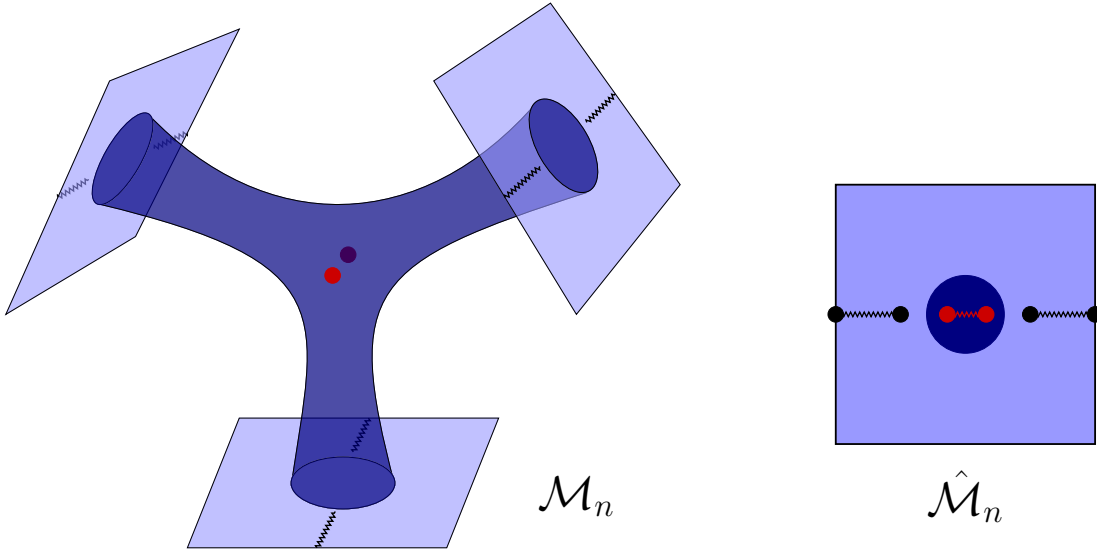


Figure 7.13: Replica wormhole solution for  $n = 3$ . The two red points are fixed points under  $\mathbb{Z}_n$  on which we insert cosmic branes and twist operators in  $\hat{\mathcal{M}}_n$ .

$$\begin{aligned} S_R^{\text{RW}} &= \lim_{n \rightarrow 1} \partial_n \left( I_G[\hat{\mathcal{M}}_n] + n \partial_n I_G[\hat{\mathcal{M}}_n] - I_G[\mathcal{M}] + \partial_n (I_{\text{CFT}}[\mathcal{M}_n] - n I_{\text{CFT}}[\mathcal{M}]) \right) \\ &= \lim_{n \rightarrow 1} (\partial_n I_G[\hat{\mathcal{M}}_n]) + \lim_{n \rightarrow 1} \partial_n (I_{\text{CFT}}[\mathcal{M}_n] - n I_{\text{CFT}}[\mathcal{M}]), \end{aligned} \quad (7.34)$$

where we have used (7.33) and RW stands for replica wormhole. The first term has exactly the same form as the only term that contributed to our calculation in section 7.5. By the same argument, this term becomes

$$\lim_{n \rightarrow 1} (\partial_n I_G[\hat{\mathcal{M}}_n]) = \frac{A}{4}, \quad (7.35)$$

where  $A$  is the area of the codimension-two surface located at the fixed point. The second term is again recognised as the replica trick formula for the entropy of the matter fields. This time, however, we have twist operators inside the gravitational region. As we discussed in section 3.3, the twist operators act as if they are glued together in the same way as the cuts are glued together. In figure 7.13 we represent this by another cut, marked in red, between the fixed points, that we should think of as being glued together with the radiation region  $R$ . Therefore the last term in (7.34) corresponds to the semiclassical entropy evaluated on  $R \cup I$ , where  $I$  is the region between the fixed points:

$$\lim_{n \rightarrow 1} \partial_n (I_{\text{CFT}}[\mathcal{M}_n] - n I_{\text{CFT}}[\mathcal{M}]) = S_{\text{semi-cl}}(\rho_{\text{CFT}}(R \cup I)). \quad (7.36)$$

Finally, to determine which of these saddles dominates the path integrals we should extremize and minimize the action. This makes the entropy equal to the minimum of the two saddles:

$$S_R = \min \left( S_R^{\text{H}}, S_R^{\text{RW}} \right). \quad (7.37)$$

This is in complete agreement with the island formula (7.4). The region  $I$  of the replica wormhole saddle then corresponds to the island region.<sup>4</sup>

<sup>4</sup>Note that in this discussion we only included the extremal cases where all the replicas were either fully disconnected or fully connected. Since we should sum over all ways of filling in the boundary conditions, we should also include the cases where only some are connected. We have ignored these because the two saddles we considered are the dominating ones [19, 20].

While this has only been a general discussion, the article by Almheiri et al. [19] also contains an explicit calculation in the above setup in which the Page curve is obtained. The calculations have the features one would expect from our general discussion: Two saddles that dominate at different times, one of them being the disconnected and the other being the fully connected replica wormhole saddle. A similar calculation was also done in a two-dimensional gravity model by Penington, Shenker, Stanford and Yang [20] using the gravitational replica trick. They also found that at early times the radiation was dominated by a disconnected saddle and at late times it was dominated by a fully connected saddle, resulting in the Page curve for the Hawking radiation.



## Chapter 8

# Summary and Discussion

We have presented what the black hole information paradox is, where it comes from, and recent insights that suggest a solution. We defined the paradox as the conflict between (1) the central dogma, which is the idea that a black hole behaves like a quantum system to an outside observer, and (2) calculations in QFT in curved spacetime, which gives Hawking radiation that is perfectly thermal. These statements cannot both be true: (1) implies the inequality  $S_{\text{rad}} \leq S_{\text{BH}}$ , which tells us that the entropy of Hawking radiation is bounded by the Bekenstein-Hawking entropy, while (2) implies that the entropy of Hawking radiation is monotonically increasing, breaking the inequality at the Page time  $t_{\text{Page}} \approx 0.54t_{\text{evap}}$ . The inequality can only break if the black hole evaporation is a non-unitary process. The Page time is within the domain of dependence of a partial Cauchy slice drawn before the black hole formation and therefore in a regime in which we would expect unitary time evolution and that the inequality would hold. On the other hand, we would also expect QFT in curved spacetime to give correct results in this regime, as the assumptions on which it is based break down only a Planck time before complete evaporation. Since it is very difficult to see why either (1) or (2) would be wrong, the black hole information paradox truly is paradoxical.

We have seen how the AdS/CFT correspondence suggests that a solution to the black hole information paradox must be one that gives the Page curve. The recently conjectured formulas for the holographic entanglement entropy provide a way in which this is possible. We performed a schematic calculation in which we showed how the quantum extremal surface formula gives the Page curve for the black hole and how the island formula gave the Page curve for the Hawking radiation. We had two surfaces in both cases: One of them was the empty surface, and the other was close to the event horizon. The empty surface gave an entropy that followed the increasing entropy of Hawking radiation as predicted by the QFT calculations. The non-empty surface gave an entropy that followed the decreasing Bekenstein-Hawking entropy. Taking the minimum of the two gave exactly the Page curve both for the black hole and for the Hawking radiation. In the case of the island formula applied to the Hawking radiation, the region between  $r = 0$  and the quantum extremal surface was the island. We discussed how entanglement wedge reconstruction helps us interpret the domain of dependence of the island as part of the radiation degrees of freedom.

Finally, we gave a general discussion on how the Page curve was derived for the entropy of Hawking radiation using the gravitational path integral with the replica trick. We defined a reduced density matrix using the gravitational path integral and replicated it. In the spirit of the gravitational path integral, we only specified the boundary condition for the geometry, leaving the rest for gravity to fill in. We found

two ways of filling in the gravitational path integral that would dominate at different times: The first was the trivial saddle, where the gravitational path integral was filled in separately for each replica. This gave a monotonically increasing entropy corresponding to the case where there is no island in the black hole. The second was the replica wormhole saddle, in which all the replicas were connected through the gravitational regions. This gave an entropy corresponding to the case where there is an island, giving an entropy that decreases along the Bekenstein-Hawking entropy. The correspondence between the replica wormhole saddle and the island formula was motivated by the simpler example in which we derived the Bekenstein-Hawking entropy through the replica trick on the Euclidean cigar. In this example, we found the entropy to be given by a quarter of the area of a codimension-two conical singularity that was present at the fixed point under the replica symmetry. This conical singularity happened to be the black hole surface, making the calculation result in the Bekenstein-Hawking entropy. In the replica wormhole saddle, we had several conical singularities in the gravitational region on which the matter fields were evaluated, corresponding to the island region in the island formula.

There are still issues with the replica wormhole calculations that need answers before we can claim to have a satisfactory solution of the black hole information paradox. First, we should point out that we have calculated the von Neumann entropy of Hawking radiation without knowing the exact density matrix  $\rho$ , as we would wish to get a complete understanding. Finding  $\rho$  seems to require the full quantum theory of gravity. It is interesting that the low-energy approach of the gravitational path integral somehow seems to know what the entropy of this state is, without knowing the exact state. Second, we have the obvious issue that the models we have considered here are not physically realistic black holes: They are black holes in two-dimensional Anti-de Sitter space glued to Minkowski spacetime on its boundary. It is not clear if this spacetime, defined as the boundary condition for the gravitational path integral, is allowed by quantum gravity. We might worry that the flat region in which there is no gravity even makes sense if this calculation is to represent effects in quantum gravity, where gravitation should be present in the entire spacetime. A more satisfactory derivation of the Page curve would be one for a physically realistic black hole formed from collapse in asymptotically flat spacetime.

We should ask why the QFT in curved spacetime calculations give what seems to be the wrong entropy of Hawking radiation. We have argued that these calculations give an entropy that does not follow the Page curve because the density matrix of Hawking radiation is exactly thermal, displaying no correlation between modes at different times. This let us set the time derivative of all the radiation emitted equal to the entropy of all modes emitted at time  $t$  as in equation (5.21), giving a monotonically increasing entropy. We may speculate that the QFT in curved spacetime calculations, which are done outside the black hole, fail to give the Page curve because they do not include quantum gravity effects inside the black hole that give correlations between early- and late-time quanta. This would have made (5.21) invalid and would allow for the entropy to not be monotonically increasing. If we interpret Hawking radiation as the result of entangled virtual particles being separated by the event horizon, the ingoing particle will eventually reach the region inside the black hole where quantum gravity effects cannot be ignored.

In chapter 4, we used the Euclidean path integral to do the equivalent of the QFT in curved spacetime calculations. The difference between this calculation and the path integral calculation in chapter 7 is that in the first case we fixed the background metric, while in the other we only fixed the boundary condition of the gravitational path integral.

This allowed replica wormholes to form, providing a way for information in the black hole interior to be escape in the Hawking radiation. It is interesting that this mechanism is very reminiscent of the ER = EPR proposal: The information escapes in the Hawking radiation because the Hawking radiation has access to the black hole interior through the replica wormholes. Yet it is unclear if we can claim the replica wormholes to be an example of ER = EPR; they are only present in the replicated, Euclidean manifold, making it difficult to see what physical significance they have. The wormholes disappear when we take the limit  $n \rightarrow 1$  at the end of the calculation. Still, they do affect the entropy, giving the Page curve.

We have discussed the hope that solving the black hole information paradox would lead to insights into quantum gravity. While we cannot claim to have a satisfactory solution yet, studying the problem has learned us more about quantum gravity. It has been important for the discoveries of the holographic principle and the ER = EPR proposal, which today are believed to be properties of quantum gravity. We are still lacking the complete theory of quantum gravity, and can hopefully learn even more from the black hole information paradox.

An interesting direction for future research would be to study how the replica wormhole saddles could be interpreted in Lorentzian spacetime, which could make its relation to ER = EPR clearer. A long-term goal would be to extend to calculations with physically realistic black holes in asymptotically flat or de Sitter spacetime. With de Sitter spacetime one could also study the radiation from the cosmological horizon. One could study if this horizon can be thought of as a quantum system in the sense of the central dogma for black holes, possibly providing insight into the quantum nature of cosmological spacetimes.

## Chapter 8. Summary and Discussion

## Appendix A

# Hypersurfaces

In this chapter we follow Poisson's description of hypersurfaces [46].

In a four-dimensional spacetime manifold, a *hypersurface* is a three-dimensional submanifold that can either be timelike, spacelike, or null. A particular hypersurface  $\Sigma$  is selected either by putting a restriction on the coordinates,

$$\Phi(x^\alpha) = 0, \quad (\text{A.1})$$

or by giving parametric equations of the form

$$x^\alpha = x^\alpha(y^a), \quad (\text{A.2})$$

where  $y^a$  ( $a = 1, 2, 3$ ) are coordinates intrinsic to the hypersurface. A unit vector  $n_\alpha$  normal to the hypersurface can be introduced if the hypersurface is not null. This is defined so that

$$n^\alpha n_\alpha = \varepsilon \equiv \begin{cases} -1 & \text{if } \Sigma \text{ is spacelike} \\ +1 & \text{if } \Sigma \text{ is timelike} \end{cases}, \quad (\text{A.3})$$

and we demand that  $n^\alpha$  point in the direction of increasing  $\Phi$ ;  $n^\alpha \Phi_{,\alpha} > 0$ . One can show that  $n_\alpha$  is given by

$$n_\alpha = \frac{\varepsilon \Phi_{,\alpha}}{|g^{\mu\nu} \Phi_{,\mu} \Phi_{,\nu}|^{1/2}} \quad (\text{A.4})$$

if the hypersurface is either spacelike or timelike. The metric intrinsic to the hypersurface  $\Sigma$  is obtained by restricting the line element to displacements confined to the hypersurface. From the parametric equations (A.2) we have that the vectors

$$e_a^\alpha = \frac{\partial x^\alpha}{\partial y^a} \quad (\text{A.5})$$

are tangent to curves contained in  $\Sigma$ . Note that this implies that  $e_a^\alpha n_\alpha = 0$ . For displacements within  $\Sigma$  we then have

$$\begin{aligned} ds_\Sigma^2 &= g_{\alpha\beta} dx^\alpha dx^\beta \\ &= g_{\alpha\beta} \left( \frac{\partial x^\alpha}{\partial y^a} dy^a \right) \left( \frac{\partial x^\beta}{\partial y^b} dy^b \right) \\ &= h_{ab} dy^a dy^b, \end{aligned} \quad (\text{A.6})$$

where

$$h_{ab} = g_{\alpha\beta} e_a^\alpha e_b^\beta \quad (\text{A.7})$$

## Appendix A. Hypersurfaces

is the *induced metric* of the hypersurface. Finally, we define the *extrinsic curvature*  $K_{ab}$  of the hypersurface as

$$K_{ab} \equiv n_{\alpha;\beta} e_a^\alpha e_b^\beta. \quad (\text{A.8})$$

The trace of the extrinsic curvature  $K$  is defined as  $K \equiv h^{ab} K_{ab}$ , where  $h^{ab}$  is the inverse of the induced metric. One can show that  $K$  is given by

$$K = n_{;\alpha}^\alpha. \quad (\text{A.9})$$

# Bibliography

- [1] S. W. Hawking. ‘Black hole explosions’. In: *Nature* 248 (1974), pp. 30–31. DOI: 10.1038/248030a0.
- [2] S. W. Hawking. ‘Particle Creation by Black Holes’. In: *Commun. Math. Phys.* 43 (1975). Ed. by G. W. Gibbons and S. W. Hawking. [Erratum: *Commun.Math.Phys.* 46, 206 (1976)], pp. 199–220. DOI: 10.1007/BF02345020.
- [3] S. W. Hawking. ‘Breakdown of Predictability in Gravitational Collapse’. In: *Phys. Rev. D* 14 (1976), pp. 2460–2473. DOI: 10.1103/PhysRevD.14.2460.
- [4] R. Penrose. *The Road to Reality: A Complete Guide to the Laws of the Universe*. Knopf Doubleday Publishing Group, 2007. ISBN: 9780679776314.
- [5] William G. Unruh and Robert M. Wald. ‘On evolution laws taking pure states to mixed states in quantum field theory’. In: *Phys. Rev. D* 52 (1995), pp. 2176–2182. DOI: 10.1103/PhysRevD.52.2176. arXiv: hep-th/9503024.
- [6] William G. Unruh and Robert M. Wald. ‘Information Loss’. In: *Rept. Prog. Phys.* 80.9 (2017), p. 092002. DOI: 10.1088/1361-6633/aa778e. arXiv: 1703.02140 [hep-th].
- [7] S. W. Hawking. ‘Black holes in general relativity’. In: *Commun. Math. Phys.* 25 (1972), pp. 152–166. DOI: 10.1007/BF01877517.
- [8] Jacob D. Bekenstein. ‘Black holes and entropy’. In: *Phys. Rev. D* 7 (1973), pp. 2333–2346. DOI: 10.1103/PhysRevD.7.2333.
- [9] G. W. Gibbons and S. W. Hawking. ‘Action integrals and partition functions in quantum gravity’. In: *Phys. Rev. D* 15 (10 May 1977), pp. 2752–2756. DOI: 10.1103/PhysRevD.15.2752. URL: <https://link.aps.org/doi/10.1103/PhysRevD.15.2752>.
- [10] Andrew Strominger and Cumrun Vafa. ‘Microscopic origin of the Bekenstein-Hawking entropy’. In: *Phys. Lett. B* 379 (1996), pp. 99–104. DOI: 10.1016/0370-2693(96)00345-0. arXiv: hep-th/9601029.
- [11] Don N. Page. ‘Average entropy of a subsystem’. In: *Phys. Rev. Lett.* 71 (1993), pp. 1291–1294. DOI: 10.1103/PhysRevLett.71.1291. arXiv: gr-qc/9305007.
- [12] Don N. Page. ‘Information in black hole radiation’. In: *Phys. Rev. Lett.* 71 (1993), pp. 3743–3746. DOI: 10.1103/PhysRevLett.71.3743. arXiv: hep-th/9306083.
- [13] Juan Martin Maldacena. ‘The Large N limit of superconformal field theories and supergravity’. In: *Adv. Theor. Math. Phys.* 2 (1998), pp. 231–252. DOI: 10.1023/A:1026654312961. arXiv: hep-th/9711200.
- [14] S. W. Hawking. ‘Information loss in black holes’. In: *Phys. Rev. D* 72 (2005), p. 084013. DOI: 10.1103/PhysRevD.72.084013. arXiv: hep-th/0507171.

## Bibliography

- [15] Netta Engelhardt and Aron C. Wall. ‘Quantum Extremal Surfaces: Holographic Entanglement Entropy beyond the Classical Regime’. In: *JHEP* 01 (2015), p. 073. DOI: 10.1007/JHEP01(2015)073. arXiv: 1408.3203 [hep-th].
- [16] Geoffrey Penington. ‘Entanglement Wedge Reconstruction and the Information Paradox’. In: *JHEP* 09 (2020), p. 002. DOI: 10.1007/JHEP09(2020)002. arXiv: 1905.08255 [hep-th].
- [17] Ahmed Almheiri et al. ‘The entropy of bulk quantum fields and the entanglement wedge of an evaporating black hole’. In: *JHEP* 12 (2019), p. 063. DOI: 10.1007/JHEP12(2019)063. arXiv: 1905.08762 [hep-th].
- [18] Ahmed Almheiri et al. ‘The Page curve of Hawking radiation from semiclassical geometry’. In: *JHEP* 03 (2020), p. 149. DOI: 10.1007/JHEP03(2020)149. arXiv: 1908.10996 [hep-th].
- [19] Ahmed Almheiri et al. ‘Replica Wormholes and the Entropy of Hawking Radiation’. In: *JHEP* 05 (2020), p. 013. DOI: 10.1007/JHEP05(2020)013. arXiv: 1911.12333 [hep-th].
- [20] Geoff Penington et al. ‘Replica wormholes and the black hole interior’. In: *JHEP* 03 (2022), p. 205. DOI: 10.1007/JHEP03(2022)205. arXiv: 1911.11977 [hep-th].
- [21] Ahmed Almheiri et al. ‘The entropy of Hawking radiation’. In: *Rev. Mod. Phys.* 93.3 (2021), p. 035002. DOI: 10.1103/RevModPhys.93.035002. arXiv: 2006.06872 [hep-th].
- [22] Murray Gell-Mann and James Hartle. ‘Quasiclassical Coarse Graining and Thermodynamic Entropy’. In: *Phys. Rev. A* 76 (2007), p. 022104. DOI: 10.1103/PhysRevA.76.022104. arXiv: quant-ph/0609190.
- [23] Sean M. Carroll. *Spacetime and Geometry*. Cambridge University Press, July 2019. ISBN: 978-0-8053-8732-2, 978-1-108-48839-6, 978-1-108-77555-7.
- [24] Izaak Neutelings. URL: <https://tikz.net/author/izaak/>. (Accessed: 13.12.2022).
- [25] Roger Penrose. ‘Gravitational collapse and space-time singularities’. In: *Phys. Rev. Lett.* 14 (1965), pp. 57–59. DOI: 10.1103/PhysRevLett.14.57.
- [26] Stephen Hawking. ‘The Occurrence of singularities in cosmology’. In: *Proc. Roy. Soc. Lond. A* 294 (1966), pp. 511–521. DOI: 10.1098/rspa.1966.0221.
- [27] Stephen Hawking. ‘The Occurrence of singularities in cosmology. II’. In: *Proc. Roy. Soc. Lond. A* 295 (1966), pp. 490–493. DOI: 10.1098/rspa.1966.0255.
- [28] Stephen Hawking. ‘The occurrence of singularities in cosmology. III. Causality and singularities’. In: *Proc. Roy. Soc. Lond. A* 300 (1967), pp. 187–201. DOI: 10.1098/rspa.1967.0164.
- [29] Cosimo Bambi. *Black Holes: A Laboratory for Testing Strong Gravity*. Springer, 2017. ISBN: 978-981-10-4524-0. DOI: 10.1007/978-981-10-4524-0.
- [30] James M. Bardeen, B. Carter and S. W. Hawking. ‘The Four laws of black hole mechanics’. In: *Commun. Math. Phys.* 31 (1973), pp. 161–170. DOI: 10.1007/BF01645742.
- [31] Robert M. Wald. *General Relativity*. Chicago, USA: Chicago Univ. Pr., 1984. DOI: 10.7208/chicago/9780226870373.001.0001.
- [32] Elizabeth Gibney. ‘How the revamped Large Hadron Collider will hunt for new physics’. In: *Nature* 605.7911 (2022), pp. 604–607. DOI: 10.1038/d41586-022-01388-6.



- [33] N. D. Birrell and P. C. W. Davies. *Quantum Fields in Curved Space*. Cambridge Monographs on Mathematical Physics. Cambridge, UK: Cambridge Univ. Press, Feb. 1984. ISBN: 978-0-521-27858-4, 978-0-521-27858-4. DOI: 10.1017/CBO9780511622632.
- [34] Robert M. Wald. ‘On Particle Creation by Black Holes’. In: *Commun. Math. Phys.* 45 (1975), pp. 9–34. DOI: 10.1007/BF01609863.
- [35] Robert M. Wald. *Quantum Field Theory in Curved Space-Time and Black Hole Thermodynamics*. Chicago Lectures in Physics. Chicago, IL: University of Chicago Press, 1995. ISBN: 978-0-226-87027-4.
- [36] Don N. Page. ‘Particle Emission Rates from a Black Hole: Massless Particles from an Uncharged, Nonrotating Hole’. In: *Phys. Rev. D* 13 (1976), pp. 198–206. DOI: 10.1103/PhysRevD.13.198.
- [37] Don N. Page. ‘Time Dependence of Hawking Radiation Entropy’. In: *JCAP* 09 (2013), p. 028. DOI: 10.1088/1475-7516/2013/09/028. arXiv: 1301.4995 [hep-th].
- [38] Joseph Schindler, Evan Frangipane and Anthony Aguirre. ‘Unitarity and the information problem in an explicit model of black hole evaporation’. In: *Class. Quant. Grav.* 38.7 (2021), p. 075025. DOI: 10.1088/1361-6382/abdf25. arXiv: 2012.07973 [gr-qc].
- [39] J. J. Sakurai and Jim Napolitano. *Modern Quantum Mechanics*. 3rd ed. Cambridge University Press, 2020. DOI: 10.1017/9781108587280.
- [40] Thomas Hartman. *Lectures on Quantum Gravity and Black Holes*. Cornell University, 2015.
- [41] Tatsuma Nishioka. ‘Entanglement entropy: holography and renormalization group’. In: *Rev. Mod. Phys.* 90.3 (2018), p. 035007. DOI: 10.1103/RevModPhys.90.035007. arXiv: 1801.10352 [hep-th].
- [42] Pasquale Calabrese and John L. Cardy. ‘Entanglement entropy and quantum field theory’. In: *J. Stat. Mech.* 0406 (2004), P06002. DOI: 10.1088/1742-5468/2004/06/P06002. arXiv: hep-th/0405152.
- [43] Bowen Chen, Bartłomiej Czech and Zi-zhi Wang. ‘Quantum information in holographic duality’. In: *Rept. Prog. Phys.* 85.4 (2022), p. 046001. DOI: 10.1088/1361-6633/ac51b5. arXiv: 2108.09188 [hep-th].
- [44] Pasquale Calabrese and John Cardy. ‘Entanglement entropy and conformal field theory’. In: *J. Phys. A* 42 (2009), p. 504005. DOI: 10.1088/1751-8113/42/50/504005. arXiv: 0905.4013 [cond-mat.stat-mech].
- [45] S. W. Hawking. ‘THE PATH INTEGRAL APPROACH TO QUANTUM GRAVITY’. In: *General Relativity: An Einstein Centenary Survey*. 1980, pp. 746–789.
- [46] Eric Poisson. *A Relativist’s Toolkit: The Mathematics of Black-Hole Mechanics*. Cambridge University Press, 2004. DOI: 10.1017/CBO9780511606601.
- [47] Luis C. B. Crispino, Atsushi Higuchi and George E. A. Matsas. ‘The Unruh effect and its applications’. In: *Rev. Mod. Phys.* 80 (2008), pp. 787–838. DOI: 10.1103/RevModPhys.80.787. arXiv: 0710.5373 [gr-qc].
- [48] Richard Tolman and Paul Ehrenfest. ‘Temperature Equilibrium in a Static Gravitational Field’. In: *Phys. Rev.* 36.12 (1930), pp. 1791–1798. DOI: 10.1103/PhysRev.36.1791.

## Bibliography

- [49] Carlo Rovelli and Matteo Smerlak. ‘Thermal time and the Tolman-Ehrenfest effect: temperature as the ‘speed of time’’. In: *Class. Quant. Grav.* 28 (2011), p. 075007. DOI: 10.1088/0264-9381/28/7/075007. arXiv: 1005.2985 [gr-qc].
- [50] Andrew Strominger. ‘Black hole entropy from near horizon microstates’. In: *JHEP* 02 (1998), p. 009. DOI: 10.1088/1126-6708/1998/02/009. arXiv: hep-th/9712251.
- [51] Don N. Page. ‘COMMENT ON ‘ENTROPY EVAPORATED BY A BLACK HOLE’’. In: *Phys. Rev. Lett.* 50 (1983), p. 1013. DOI: 10.1103/PhysRevLett.50.1013.
- [52] W.V Quine. *The Ways of Paradox and Other Essays*. eng. Cambridge, Mass, 1976.
- [53] Daniel Harlow. ‘Jerusalem Lectures on Black Holes and Quantum Information’. In: *Rev. Mod. Phys.* 88 (2016), p. 015002. DOI: 10.1103/RevModPhys.88.015002. arXiv: 1409.1231 [hep-th].
- [54] L. Susskind and J. Lindesay. *An introduction to black holes, information and the string theory revolution: The holographic universe*. 2005.
- [55] Gerard ’t Hooft. ‘Dimensional reduction in quantum gravity’. In: *Conf. Proc. C* 930308 (1993), pp. 284–296. arXiv: gr-qc/9310026.
- [56] Leonard Susskind. ‘The World as a hologram’. In: *J. Math. Phys.* 36 (1995), pp. 6377–6396. DOI: 10.1063/1.531249. arXiv: hep-th/9409089.
- [57] David Tong. ‘String Theory’. In: (Jan. 2009). arXiv: 0908.0333 [hep-th].
- [58] Horatiu Nastase. *Introduction to the ADS/CFT Correspondence*. Cambridge University Press, Sept. 2015. ISBN: 978-1-107-08585-5, 978-1-316-35530-5.
- [59] Juan Martin Maldacena. ‘Eternal black holes in anti-de Sitter’. In: *JHEP* 04 (2003), p. 021. DOI: 10.1088/1126-6708/2003/04/021. arXiv: hep-th/0106112.
- [60] Juan Maldacena and Leonard Susskind. ‘Cool horizons for entangled black holes’. In: *Fortsch. Phys.* 61 (2013), pp. 781–811. DOI: 10.1002/prop.201300020. arXiv: 1306.0533 [hep-th].
- [61] Albert Einstein and N. Rosen. ‘The Particle Problem in the General Theory of Relativity’. In: *Phys. Rev.* 48 (1935), pp. 73–77. DOI: 10.1103/PhysRev.48.73.
- [62] Albert Einstein, Boris Podolsky and Nathan Rosen. ‘Can quantum mechanical description of physical reality be considered complete?’ In: *Phys. Rev.* 47 (1935), pp. 777–780. DOI: 10.1103/PhysRev.47.777.
- [63] Shinsei Ryu and Tadashi Takayanagi. ‘Holographic derivation of entanglement entropy from AdS/CFT’. In: *Phys. Rev. Lett.* 96 (2006), p. 181602. DOI: 10.1103/PhysRevLett.96.181602. arXiv: hep-th/0603001.
- [64] Veronika E. Hubeny, Mukund Rangamani and Tadashi Takayanagi. ‘A Covariant holographic entanglement entropy proposal’. In: *JHEP* 07 (2007), p. 062. DOI: 10.1088/1126-6708/2007/07/062. arXiv: 0705.0016 [hep-th].
- [65] Bowen Chen, Bartłomiej Czech and Zi-zhi Wang. ‘Quantum information in holographic duality’. In: *Rept. Prog. Phys.* 85.4 (2022), p. 046001. DOI: 10.1088/1361-6633/ac51b5. arXiv: 2108.09188 [hep-th].
- [66] Aitor Lewkowycz and Juan Maldacena. ‘Generalized gravitational entropy’. In: *JHEP* 08 (2013), p. 090. DOI: 10.1007/JHEP08(2013)090. arXiv: 1304.4926 [hep-th].

- [67] Mukund Rangamani and Tadashi Takayanagi. *Holographic Entanglement Entropy*. Vol. 931. Springer, 2017. DOI: [10.1007/978-3-319-52573-0](https://doi.org/10.1007/978-3-319-52573-0). arXiv: [1609.01287](https://arxiv.org/abs/1609.01287) [hep-th].