LU-TP 21-29
June 2021

# Recent Progress on the Black Hole Information Paradox

**Eivind Jørstad**

Department of Astronomy and Theoretical Physics, Lund University

Master thesis supervised by Dr. Michal P. Heller and Dr. Stefan Prestel


LUND
UNIVERSITY

# Abstract

The black hole information paradox is of central importance in the study of quantum gravity. The paradox states that quantum fields surrounding a black hole behave in ways that run into conflict with the unitarity of quantum mechanical time evolution. In the course of the past two years, new theoretical discoveries have made progress towards the paradox's resolution, suggesting that unitarity can be saved. The central result is that a generalized entanglement entropy, originally arising in the anti-de Sitter/conformal field theory correspondence, gives entanglement dynamics for the quantum fields that is consistent with unitarity. In this thesis, we review the black hole information paradox and the recent results. In addition, we explore a puzzling feature of the generalized entropy and show a novel result regarding the entanglement entropy in the interior of an evaporating black hole in the Russo-Susskind-Thorlacius model. We also apply an approximation scheme from condensed matter physics in a novel way, using a quasiparticle description to compute entanglement entropy in a moving-mirror model of black hole evaporation.

# Acknowledgments

I would like to thank my supervisor Dr. Michal P. Heller, whose advice, enthusiasm and knowledge has been invaluable during all stages of the project. I am also very grateful to Dr. Ignacio A. Reyes and Dr. Pedro F. Ramirez, who have both provided helpful discussions throughout the project's duration. Finally, I would like to thank Dr. Stefan Prestel for acting as my local supervisor during this project. The group *Gravity, Quantum Fields and Information* at the Max Planck Institute for Gravitational Physics (Albert Einstein Institute), where this thesis was prepared, is supported by the Alexander von Humboldt Foundation and the Federal Ministry for Education and Research through the Sofja Kovalevskaja Award.

# Popular description

Black holes are arguably the most extreme objects in the universe. The strength of their gravitational attraction is such that if you get too close – and here "you" can refer to anything from a human in a spaceship to a single particle of light – you will be pulled in towards its center without any possibility of escape. Due to the intensity of these gravitational forces, physicists have often wondered whether one can make sense of quantum mechanics – our best description physics at very small length-scales – in the presence of black holes. In other words, one can ask: how does quantum mechanical matter behave around black holes, and can a black hole itself be considered a quantum system?

In 1975, the physicist Stephen Hawking published a now famous calculation, in which he looked at quantum mechanical matter near the surface of a black hole, called its *event horizon*. What he found is that the black hole should not be completely black; quantum effects will produce a small amount of radiation which causes the black hole to evaporate. It was soon realized that this has significant consequences. In particular, Hawking noted that if the black hole evaporates away completely then quantum mechanics appears to be violated. In other words, the evolution process that starts with a black hole and ends with only a cloud of radiation is not allowed by the equations of quantum mechanics. This apparent contradiction was named the *black hole information paradox*.

Almost 20 years later, in 1993, the information paradox was sharpened by Don Page, who realized that black hole evaporation leads to trouble even before the black hole evaporates away completely. The issue has to do with *quantum entanglement*, a curious feature of quantum mechanics in which properties of quantum systems can be correlated even though the systems are located far apart from each other. In particular, throughout the evaporation process the amount of entanglement between the black hole and the radiation increases steadily, while at the same time the black hole's capacity for entanglement decreases due to energy loss in the form of radiation. The conflict arises around the so-called *Page time*, where the radiation appears to grow more entangled with the black hole than what should be possible according to quantum mechanics.

Finally, in 2019 it was discovered that Page's issue might be avoided in a surprising manner. A new method for computing the entropy of the black hole was conjectured based on a deep connection between gravity and certain quantum theories, and the resulting entropy narrowly avoids Page's paradox. Moreover, it does so by showing the exact behaviour Page predicted for quantum mechanical black holes.

In this thesis we review these recent developments, and discuss a puzzling result of the conjectured entropy formula. In addition, we show that a certain useful approximation scheme known from condensed matter physics can be applied to a class of models describing black hole evaporation.

# Contents


**I Black Holes, Entropy and Hawking Radiation** **6**

**1 Introduction** **6**

**2 Black Hole Thermodynamics** **7**

**3 Hawking Radiation** **10**

**4 Radiation from Moving Mirrors** **12**

**5 A Thermodynamical Paradox** **15**

**6 Entropy in QFT and Gravity** **17**

6.1 The Replica Trick . . . . . 18

6.2 Entropy in the AdS/CFT correspondence: . . . . . 21

6.3 AdS/CFT and the Replica Trick . . . . . 24

**II Quantum Extremal Islands** **26**

**7 Generalized Entropy** **26**

7.1 Islands in the RST model . . . . . 28

7.1.1 Review of the RST model . . . . . 28

7.1.2 The Page curve . . . . . 34

**8 A Puzzle Regarding Islands** **36**

8.1 Violating the Bekenstein-Hawking Bound in the RST Model . . . . . 37

**III Moving Mirrors and Quasiparticles** **39**

**9 The Alba-Calabrese Formula** **39**

**10 Quasiparticle Picture for Moving Mirrors** **40**

**IV Conclusions** **44**

**11 Summary and Outlook** **44**

**A Conformal Field Theory** **45**

A.1 Toolbox . . . . . 46

**B Penrose Diagrams** **47**

# Part I

# Black Holes, Entropy and Hawking Radiation

## 1 Introduction

The theory of general relativity is an extremely successful description of gravity. Interestingly, it is a purely classical field theory, which differentiates it from the quantum field theoretic descriptions of all the other fundamental forces. Naive attempts at quantization run into difficulties, as the resulting theory can be showed to be non-renormalizable. In practice, the comparative weakness of gravitational interaction means that its effect can be neglected in most situations where quantum field theory calculations become relevant.

However, black holes provide a setting where gravitational effects are predicted to give rise to novel quantum phenomena. In particular, Hawking predicted [1] that an inertial observer asymptotically far away from a black hole will detect out-going thermal radiation. The central idea behind the derivation is that in curved spacetimes, inertial observers are no longer related by simple Lorentz transformations. As a consequence, the vacuum states defined by different inertial observers can be inequivalent, and they therefore do not agree on a notion of particle number. In particular, the vacuum of an inertial observer near the horizon corresponds to a thermal spectrum at asymptotic infinity.

Since its publication, Hawking's prediction has played an important role in quantum gravity research. A central question has been whether the black hole evaporation process is compatible with the unitarity of quantum mechanical time evolution. Hawking suggested that is was not, essentially due to the mixed nature of the thermal state of the radiation [2]. A particular instance of this question was raised by Page [3, 4], who argued that the thermodynamic Bekenstein-Hawking entropy does not appear to originate from a quantum theory of gravity, as the the entanglement entropy of Hawking radiation would exceed the limits placed by unitarity.

Page's issue has recently received much attention due to the surprising results of [5, 6], in which entanglement entropy formulas motivated by the anti-de Sitter/conformal field theory (AdS/CFT) correspondence are applied to evaporating black holes. Surprisingly, it is found that these expressions only agree with the usual semi-classical methods up to the point where unitarity would be violated, at which point the behaviour changes drastically and unitarity is saved.

This thesis is divided into three parts. Part I contains the necessary background material needed to discuss the results of the previous paragraphs. We first introduce black hole thermodynamics and Hawking radiation, which lead us to consider the Page curve for

entanglement entropy. We also introduce the AdS/CFT correspondence, focusing on its relation to entanglement entropy. These are all vast topics of research, which could be the topic of a thesis in their own right. As such, we focus on the central results needed to in the later parts of the thesis. Part II is then devoted to a discussion of quantum extremal islands. We study the appearance of islands for an evaporating black hole in the Russo-Susskind-Thorlacius (RST) model, which is a two dimensional dilaton theory of gravity coupled to matter. In the semi-classical limit one can include the back-reaction of quantum fields on the classical geometry, giving an evaporating black hole that shrinks away due to Hawking-radiation. We also discuss a puzzling feature of the generalized entropy formula, which to our knowledge has not been emphasized in the literature. Finally, in Part III we show that a well known approximation from the study of quantum lattice systems can be applied to moving-mirror models. We also take the first steps toward applying this approximation to the RST-model. Throughout the thesis we work in natural units, such that $\hbar = c = k_B = 1$.

# 2 Black Hole Thermodynamics

*Black Holes:*

In 1916, Schwarzschild discovered the first exact non-trivial solution to Einstein's field equations. The solution is known as the Schwarzschild metric,

$$ds^2 = -\left(1 - \frac{2G_N M}{r}\right) dt^2 + \left(1 - \frac{2G_N M}{r}\right)^{-1} dr^2 + r^2 d\Omega^2, \tag{2.1}$$

and is the unique vacuum-solution with a spherically symmetric and static metric [7]. Here, $G_N$ is Newtons constant. The solution is a good approximation of the spacetime outside a static non-rotating star, and is therefore of physical interest.

The metric eq. (2.1) has two particularly interesting properties. Firstly, at the origin of coordinates, we get a divergence in the coordinate invariant quantity

$$R_{\alpha\beta\mu\nu} R^{\alpha\beta\mu\nu} = \frac{48(G_N M)^2}{r^6}, \tag{2.2}$$

where $R_{\alpha\beta\mu\nu}$ is the usual Riemann tensor, showing that the geometry is singular at this point. Secondly, the $r$-direction becomes timelike for $r < 2G_N M$. In fact, all non-spacelike curves in this region terminate at the singularity, meaning that no physical signal can escape out to $r > 2G_N M$. This is the reason for calling eq. (2.1) a ***black hole***; a physical trajectory on the inside will never escape out to the exterior. The surface $r = 2G_N M$ is known as the ***event horizon***.

The maximal extension of the Schwarzschild geometry[1] is illustrated by the Penrose diagram (see appendix B) to the left in Fig. 1. This geometry has a number of peculiar

[1] That is, the geodesically complete manifold whose metric is given everywhere by eq. (2.1) up to coordinate transformations.

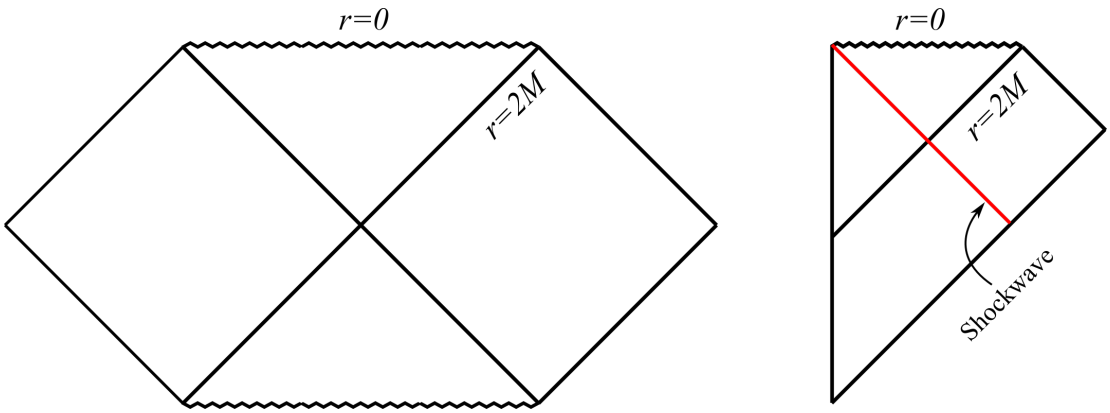


Figure 1: ***Left:*** A Penrose diagram showing the causal structure of an maximally extended Schwarzschild geometry. There is a past and future singularity, as well as two disconnected exterior regions. ***Right:*** A Schwarzschild black hole forming from the collapse of a lightlike shock wave. In this case the spacetime region to the future of the shock wave is described by the Schwarzschild metric. For more details see appendix B.

features, such as the presence of both past and future singularities. One can also have a spacetime where eq. (2.1) only describes a patch. The right hand side of Fig.1 shows such an example, where a black hole forms from a collapsing shell of matter (referred to as a shock wave). In this case, the metric to the future of the shock wave is given by eq. (2.1).

The Schwarzschild metric is just a particularly simple example of a black hole geometry. There are several others, including black holes with angular momentum and electrical charge. Some of these geometries can be realized in nature, e.g. after the collapse of a massive star, and we are now at a point where these can be detected directly using gravitational wave astronomy or radio telescopes.

*Thermodynamics:*

Around 50 years ago, several results were obtained that indicated a connection between black hole physics and thermodynamics. For an overview of these results, see [8]. One example is the first law of black hole mechanics, which describes a black hole in equilibrium subject to small perturbations. It takes the form

$$dE = \frac{\kappa}{8\pi} dA + \Omega dJ + \Phi dQ, \tag{2.3}$$

where $E$ is the black hole energy, $A$ is the area of the horizon, $J$ is its angular momentum, $\Omega$ is the angular velocity at the horizon[2], $Q$ is the total charge and $\phi$ is the electric potential at the horizon. $\kappa$ is called the surface gravity and is given by

$$\kappa = (4MG_N)^{-1} \tag{2.4}$$

[2]More specifically, $\Omega$ is the angular velocity of an orbiting inertial observer with zero angular momentum near the horizon, which can be non-zero due to frame-dragging. More details can be found in [7].

for the Schwarzschild black hole[3]. Importantly, for a stationary black hole the coefficients in eq. (2.3) are constant on the black hole horizon [8]. The last two terms represent work done on the black hole, so if one can make the identification $\frac{\kappa}{8\pi}dA \sim TdS$, one would obtain the first law of thermodynamics $dE = TdS + dW$.

The identification of $A$ as an entropy is further supported by the area theorem, which states that the horizon area obeys

$$\Delta A \geq 0, \tag{2.5}$$

for a general class of processes. The relation is for example obeyed by well known processes such as the extraction of energy from a black hole via the Penrose process and for the merging of two black holes. This gives a second law of thermodynamics for black holes, with $A$ playing the role of entropy [9]. Reversible processes leave the area unchanged, in agreement with the thermodynamical interpretation.

One problem with the analogy between black holes and thermodynamic systems is that classical stationary black holes do not give off energy, and therefore do not have a physical temperature. However, this changes when quantum effects are accounted for. In fact, Hawking radiation (see below) gives the black hole a temperature that depends on $\kappa$ in exactly the right way to make the identification $dS \sim dA$. Using $T_H = \frac{\kappa}{2\pi}$ (eq. (3.7)), we get the resulting entropy

$$S_{BH} = \frac{A}{4G_N}, \tag{2.6}$$

which is known as the Bekenstein-Hawking entropy. Interestingly, it is the inclusion of a quantum effect that allows us to give the black hole a proper thermodynamic interpretation. Note that the Bekenstein-Hawking entropy scales with area, as opposed to the entropy of regular matter (e.g. an ideal gas) which scales with volume. Furthermore, black holes in GR are completely described by their mass, angular momentum and charge, having no microscopic structure. It is then interesting to ask whether $S_{BH}$ counts the microstates of a more fundamental theory of gravity. That is, one can ask if there is a relation on the form

$$S_{BH} \sim \ln N, \tag{2.7}$$

where $N$ is the number of black hole microstates compatible with given macroscopic observables. There are suggestive results in this direction from string theory [10], where one can obtain the Bekenstein-Hawking entropy by counting the microstates of very specific black holes. Similarly, the Bekenstein-Hawking entropy has also been derived in loop quantum gravity, see e.g. [11]. We will return to this question in section 5.

[3]General covariant definitions of $\kappa$ and the other quantities in eq. (2.3) can be found in [7].

# 3 Hawking Radiation

In 1975, Hawking predicted that black holes radiate particles with a thermal blackbody spectrum [1]. This counterintuitive result is interesting for several reasons, particularly because it gives rise to a series of paradoxes. Here we aim to give an overview of the key results, while a more quantitative treatment in the RST-model is given in section 7.1. For a pedagogical introduction to Hawking radiation we refer the reader to [12] and [13].

Hawking radiation is derived in the *semi-classical* limit of quantum gravity, where all matter fields are quantized while spacetime is kept classical. Recall first conventional quantum field theory, where quantization is carried out by promoting field values to operators and imposing appropriate commutation relations, as seen in any standard textbook (such as [14]). One can choose a complete set of positive frequency modes $\phi_k$ which satisfy the classical equations of motion[4], and expand the field operator into positive and negative frequency modes

$$\Phi = \int dk \left( a_k \phi_k + a_k^\dagger \phi_k^* \right), \tag{3.1}$$

where we have chosen the example of a scalar field $\Phi$, and $k$ is a label specifying the modes. The operators $a_k^\dagger$ and $a_k$ are creation and annihilation operators associated with the mode $\phi_k$. In flat spacetime one typically chooses the modes to be positive frequency plane waves $\phi_{\bar{p}} \sim e^{-i(\omega t - \bar{p}\cdot\bar{x})}$, where $(\omega, \bar{p})$ is a four-momentum and where $\omega = \sqrt{\bar{p}^2 + m^2}$. One reason is that in this case Lorentz-transformations map the positive frequency modes to positive frequency modes[5], which implies that the Fock vacuum (i.e. the state satisfying $a_{\bar{p}} |0\rangle = 0$ for all choices of $\bar{p}$) is the same state for any choice of inertial reference frame. That is to say, in flat spacetime all inertial observers can agree on a choice of vacuum state.

In curved space, the above discussion is complicated by the fact that each inertial observer has their own locally inertial frame, with no pre-ordained relation between the frames of different observers. This prohibits the definition of an observer independent vacuum state[6]. More precisely, consider again the scalar field $\Phi$, and suppose two observers are using different sets of positive frequency modes $\phi_k$ and $\psi_k$. By virtue of being complete sets of modes, we know that any mode in the first set can be expanded in terms of the other set as

$$\phi_k = \int dl \left( \alpha_{lk} \psi_l + \beta_{lk}^* \psi_l^* \right), \tag{3.2}$$

where $\alpha_{lk}$ and $\beta_{lk}$ are complex valued coefficients. Inserting eq. (3.2) into eq. (3.1), we get

[4]To be more precise, the modes $\phi_k$ must be such that any solution to the equations of motion can be written as a linear combination of $\phi_k$'s and $\phi_k^*$'s. They must also satisfy technical conditions that ensure the corresponding creation and annihilation operators have correct commutation relations [13]

[5]That is: $\omega \to \omega' = \gamma(\omega - \beta p_i)$, for a boost in the $x_i$ direction. Since $\omega = \sqrt{m^2 + \bar{p}^2} > |p_i|$, and $\beta < 1$, we get $\omega' > 0$.

[6]In fact this is true even in flat space if one considers non-inertial observers, as shown by the Unruh effect. For an introduction to the Unruh effect, see [15].

the expansion of $\Phi$ in terms of $\psi_l$:

$$\Phi = \int dl \left( a'_l \psi_l + a'^{\dagger}_l \psi^*_l \right), \tag{3.3}$$

where

$$a'_l = \int dk \left( \alpha_{lk} a_k + \beta_{lk} a^{\dagger}_k \right). \tag{3.4}$$

The coefficients $\alpha_{lk}$ and $\beta_{lk}$ are known as Bogoliubov coefficients. Now, let us suppose we are in the Fock vacuum defined by the $\phi_k$ modes, and let evaluate the expectation value of the number operator $N(\psi_l) = a'^{\dagger}_l a'_l$ for the $\psi_l$ mode. Using eq. (3.4), we get

$$\langle N(\psi_l) \rangle = \int dk \, |\beta_{lk}|^2. \tag{3.5}$$

Thus, the vacua defined by to different observers are not equivalent if there are non-zero $\beta$-coefficients that relate their sets of modes. In this case, the observers will generally disagree on the notion of particle number.

In [1], Hawking considers two observers around a Schwarzschild black hole formed from collapse. One observer is placed asymptotically far away from the black hole, where space-time is approximately flat, and one is infalling inertial observer close to the event horizon. A wave packet $\psi$ in the asymptotic flat region, containing only positive frequency modes according to the asymptotic observer, can then be propagated backwards in time until it reaches the infalling observer near the horizon. Expanding the wave packet in the modes of the infalling observer, it turns out that one must include both positive and negative frequencies. Thus, there must be non-zero $\beta$-coefficients relating the two sets of modes, and as a consequence the vacua of the two observers are not equivalent.

Now, assuming we are in the vacuum state defined by the infalling observer, we see from eq. (3.5) that the asymptotic observer measures a non-zero amount of particles[7]. At late times, for a wave packet $\psi$ sharply peaked around the frequency $\omega$, Hawking found [1]

$$\langle N(\psi) \rangle = \frac{\Gamma(\psi)}{e^{\frac{2\pi\omega}{\kappa}} - 1}, \tag{3.6}$$

where $\kappa$ is the surface gravity. This is the spectrum of a radiating black body, with an additional factor $\Gamma$, which can be interpreted as the probability of the wave packet escaping to infinity instead of begin absorbed into the black hole. The temperature defined by this spectrum is called the Hawking temperature and is given by

$$T_H = \frac{\kappa}{2\pi}, \tag{3.7}$$

[7] One should really be careful here, and argue that this constitutes physical particles that can be observed by a detector. For such considerations, see [16].

where $\kappa$ is the surface gravity, given by eq. (2.4) for a Schwarzschild black hole. As we saw above, the existence of a physical black hole temperature allows us to interpret eq. (2.3) and eq. (2.5) as proper thermodynamic relations.

One should also point out an important consequence of the black hole horizon. Specifically, to describe the radiation which is located outside the horizon, we have to trace out degrees of freedom in the interior. As a result, we might get a mixed state, described by a density matrix instead of state vector. In fact, for Hawking radiation the density matrix turns out to be that of a thermal ensemble at the Hawking temperature, in accordance with the above result.

# 4 Radiation from Moving Mirrors

Although Hawking radiation arises in the context of gravity, we can demonstrate a similar effect in flat space. The idea is to introduce a moving boundary on which we impose reflecting boundary conditions for the fields. This gives two natural choices of modes which have nonequivalent vacua, and as a result we get particle creation in the manner described around eq. (3.5). These *moving mirror models* have been studied extensively as toy-models for Hawking radiation, as they do not suffer from some of the complications that arise in curved space, for a recent example see [17]. In fact, we will later discuss a black hole toy-model which is closely related to moving mirrors, see section 7.1.

To start, consider a quantum field theory living on two-dimensional Minkowski space with light cone coordinates

$$u \equiv t - x, \qquad v \equiv t + x, \tag{4.1}$$

and let us first focus on a stationary mirror defined by the curve $v = u$ (i.e. $x = 0$). One way to impose reflecting boundary conditions on the mirror is to quantize the theory in terms of modes that obey the boundary condition. Take for example the massless free scalar, for which the boundary condition on the mirror is just the Dirichlet condition. A natural choice of modes that fulfills this condition is the linear combination of (complex) plane waves that is identically zero on the mirror [18]:

$$\phi_\omega = \frac{1}{\sqrt{4\pi\omega}} \left( e^{-i\omega v} - e^{-i\omega u} \right), \tag{4.2}$$

where we label the modes by the positive frequency $\omega$. These modes are classical standing waves. The set of modes is complete, and satisfies the equation of motion $\partial_u \partial_v \phi_\omega = 0$. We get creation and annihilation operators $a_\omega^\dagger$ and $a_\omega$, which defines our vacuum state $a_\omega |0\rangle = 0$ for all $\omega$. For brevity, we will here assume that this vacuum state contains no particles or energy flux, as is discussed in more detail in [16]. We may then define the trajectory of a moving mirror as the implicit curve $v = p(u)$, for some choice of the

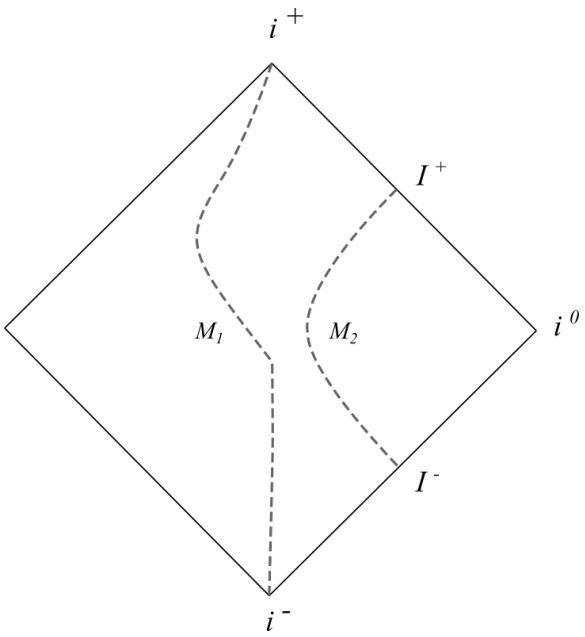


Figure 2: A Penrose diagram of Minkowski space showing two examples of mirror trajectories $M_1$ and $M_2$. The trajectory $M_1$ starts of as stationary, and then moves in a timelike fashion. The trajectory $M_2$ is asymptotically lightlike, and therefore touches the boundary at null infinity $I^{\pm}$. We consider fields living to the right a given mirror trajectory.

function $p$. This gives two choices of modes that are analogous to eq. (4.2):

$$\phi_\omega = \frac{1}{\sqrt{4\pi\omega}} \left( e^{-i\omega v} - e^{-i\omega p(u)} \right), \tag{4.3}$$

$$\psi_\omega = \frac{1}{\sqrt{4\pi\omega}} \left( e^{-i\omega p^{-1}(v)} - e^{-i\omega u} \right), \tag{4.4}$$

where both vanish on the mirror trajectory by definition (note, $p^{-1}$ is the inverse function of $p$ and not its reciprocal). These modes are referred to as in- and out-modes respectively. We see that the in-mode has the form of an incoming plane wave which is reflected by the mirror in a distorted way characterized by the function $p(u)$, and similarly the out-mode describes an incoming wave characterized by $p^{-1}(u)$ which is reflected as an outgoing plane wave. Note that if we have a mirror trajectory like $M_1$ in Fig. 2, which is stationary up until some time $t_0$, then these modes agree with eq. (4.2) up until the mirror moves.

To calculate the Bogoliubov coefficients, one uses the fact that that each set of modes is orthonormal under the following inner product

$$(f, g) \equiv -i \int dx \, \left( f \partial_t g^* - g^* \partial_t f \right), \tag{4.5}$$

as is readily confirmed. Indeed, we have $(\phi_\omega, \phi_{\omega'}) = (\phi^*_\omega, \phi^*_{\omega'}) = \delta(\omega - \omega')$, and $(\phi_\omega, \phi^*_{\omega'}) = 0$,

and similarly for the out-modes.[8] Looking back at eq. (3.2), we then find:

$$\alpha_{ww'} = (\phi_\omega, \psi_{\omega'}), \tag{4.6}$$

$$\beta_{ww'} = -(\phi_\omega, \psi^*_{\omega'}), \tag{4.7}$$

which can be used to obtain the expectation value of the number operator, as in eq. (3.5).

It might also be interesting to check whether the moving mirror gives rise to an energy flux. To this end, we can exploit the fact that the free massless scalar has conformal symmetry (see appendix A). This means that the action is invariant under angle-preserving maps, and as a useful consequence we can determine the behaviour of the stress energy tensor under these transformations. In particular, consider the transformation

$$u \to p(u), \quad v \to v, \tag{4.8}$$

which maps the mirror to a static trajectory. The in-mode vacuum corresponds to the static-mirror vacuum in these new coordinates. Since the vacuum state for the static mirror does not contain any energy-flux, the transformation law eq. (A.16) immediately gives us[9]

$$\langle T_{uu} \rangle = -\frac{1}{24\pi}\{p(u), u\} = -\frac{1}{24\pi}\left(\frac{p'''(u)}{p'(u)} - \frac{3}{2}\left(\frac{p''(u)}{p'(u)}\right)^2\right). \tag{4.9}$$

The same transformation law tells us that the other components of the stress-energy remain at zero. We see there is a non-zero outgoing flux of energy, provided the function $p$ does not satisfy $\{p(u), u\} = 0$. One might note eq. (4.9) is not necessarily positive, which is also a general feature of the stress-energy tensor in curved space [16].

*Example:*

As a demonstration, we may consider the mirror trajectory defined by

$$p^{-1}(v) = -\kappa^{-1} \log\left(\frac{B - v}{B}\right), \tag{4.10}$$

where $\kappa$ and $B$ are real constants. This well known trajectory has the property that it produces thermal radiation [16, 19]. Indeed, integrating eq. (4.7) by parts, we get the expression

$$\beta_{\omega'\omega} = -(2\pi)^{-1}\left(\frac{\omega}{\omega'}\right)^{\frac{1}{2}} \int_{-\infty}^{B} dv\, e^{-i\omega v - i\omega' p^{-1}(v)}, \tag{4.11}$$

which can be integrated to yield $\beta_{\omega'\omega} \sim \Gamma(1 + i\omega'\kappa^{-1})$. We then find[10]

$$|\beta_{\omega'\omega}|^2 = \frac{1}{2\pi\kappa\omega}\frac{1}{e^{2\pi\omega'\kappa} - 1}, \tag{4.12}$$

[8] Note also that we can evaluate the inner product at any time $t$, as the time derivative of $(f, g)$ vanishes for any choice of $f$ and $g$ that both satisfy the equations of motion.

[9] Using the fact that $c = 1$ for a free massless scalar.

[10] Using the identity $|\Gamma(1 + ia)|^2 = \pi a / \sinh(\pi a)$, where $a$ is a real number.

which looks like a thermal spectrum at temperature $T = \kappa/2\pi$ (compare with eq. (3.6)). Note that inserting this into eq. (3.5) yields a logarithmic divergence which can be attributed to the fact that the mirror is continuously producing radiation at all times, creating an infinite number of particles per mode [16]. This complication does not arise if the mirror has uniform velocity at asymptotically early times. Using eq. (4.9), we also find

$$\langle T_{uu} \rangle = \frac{\kappa^2}{48\pi}, \tag{4.13}$$

showing that there is a positive outgoing energy flux.

# 5 A Thermodynamical Paradox

In this section we discuss important consequences of Hawking radiation. First, we briefly outline the original information paradox, which is concerned with the late stages of black hole evaporation. Then, we introduce a thermodynamic paradox raised by Page [3, 4], which relates to the discussion of black hole entropy in section 2. The latter issue plays a central role in Part II of the thesis.

*Hawking's information paradox:*

The results of the previous section run into problems with the unitarity of quantum mechanical time evolution if they are extrapolated to the late stages of evaporation, as noted by Hawking in [2]. The unitarity of the time evolution operator $\mathcal{U}$ is an important feature of quantum mechanics, as it implies the conservation of probability[11],

$$\mathrm{Tr}\left[\rho(t_1)\right] = \mathrm{Tr}\left[\mathcal{U}(\Delta t)\rho(t_0)\mathcal{U}^\dagger(\Delta t)\right] = \mathrm{Tr}\left[\rho(t_0)\right], \tag{5.1}$$

which follows from the cyclicity of the trace, while also implying the reversibility of time evolution,

$$\rho(t_0) \neq \rho'(t_0) \Longleftrightarrow \rho(t_1) \neq \rho'(t_1), \tag{5.2}$$

which simply follows from the the fact that the unitary operators are invertible. Since the radiation is described by a purely thermal density matrix $\rho_{\mathrm{thermal}}$, the radiation state is independent of the state of any matter which could have fallen into the black hole at an earlier time. Thus, a black hole evaporating down to zero size (leaving only radiation) would allow evolution on the form,

$$\rho_{\mathrm{initial}} \rightarrow \rho_{\mathrm{thermal}} \leftarrow \rho'_{\mathrm{initial}}, \tag{5.3}$$

where $\rho_{\mathrm{initial}} \neq \rho'_{\mathrm{initial}}$. In that case, unitarity would be violated since arbitrary infalling matter states all evolve into the same thermal state of radiation. However, one is not forced to conclude that unitarity is violated, since the semi-classical approximation is expected to

[11] For pure states, this gives the perhaps more familiar statement $|\langle\psi(t_1)|\psi(t_1)\rangle|^2 = |\langle\psi(t_0)|\psi(t_0)\rangle|^2$.

break down at late times [20]. Thus, Hawking's results are untrustworthy in this regime, and for this reason we will not be focusing in this version of the paradox in this thesis. More details about this version of the paradox can be found in [12].

*Page's Paradox:*

In this thesis, we will focus on a unitarity-related issue raised by Page [3, 4], which occurs before the breakdown of the semi-classical approximation. The issue relates to entanglement entropy, which is defined by

$$S(\rho) = -\mathrm{Tr}\left(\rho \ln \rho\right), \tag{5.4}$$

where $\rho$ is a density matrix describing some quantum system[12]. If we only wish to describe a subsystem $A$ and ignore its complement $\bar{A}$, we can obtain a density matrix for $A$ by taking the partial trace[13]

$$\rho_A = \mathrm{Tr}_{\bar{A}}\left(\rho\right). \tag{5.5}$$

For a system in a pure state (i.e. $\rho = |\psi\rangle\langle\psi|$ for some state-vector $|\psi\rangle$) one can see that $S(\rho) = 0$, which implies $S(\rho_A) = S(\rho_{\bar{A}})$. If the Hilbert space of the system $A$ is finite dimensional, as it would be for a spin-chain representation of quantum field theory, we also have

$$0 \leq S(\rho_A) \leq \ln\left[\dim\left(A\right)\right], \tag{5.6}$$

which is straightforward to prove using induction. The term "entanglement entropy" derives from the fact that the lower bound is reached if (and only if) $A$ and $\bar{A}$ are uncorrelated, while the upper bound saturates if (and only if) the two systems are maximally entangled.

One can now consider the entanglement entropy of Hawking radiation, and test whether its behaviour is consistent with a thermodynamical interpretation of the Bekenstein-Hawking entropy. In particular, let us assume for the moment that the Bekenstein-Hawking entropy counts the microstates of a more fundamental quantum mechanical theory of gravity. Fixing the set of macroscopic observables $A_i$ that characterize a classical black hole, we are interested in the set of states $\rho_{bh}$ that gives the correct expectation values

$$A_i = \langle \hat{A}_i \rangle = \mathrm{Tr}\left(\rho_{bh}\hat{A}_i\right). \tag{5.7}$$

One can then define a quantum thermodynamic entropy, which satisfies the second law [22],

$$S_{\mathrm{th}} = \underset{\rho_{\mathrm{bh}}}{\mathrm{Max}}\left[S(\rho_{\mathrm{bh}})\right], \tag{5.8}$$

where the maximization runs over all states compatible with the macroscopic observables. This definition is not so unfamiliar as it first seems; if the Hilbert sub-space consistent with

[12]Recall that a density matrix is an object that can represent both pure and mixed quantum states. In general, it takes the form $\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|$, where $p_i$ is a real number, for an appropriate choice of basis states.

[13]Defining the partial trace in quantum field theory is a subtle issue, see for example [21]. In section 6.1, we show how one can take the partial trace for a free massless scalar.

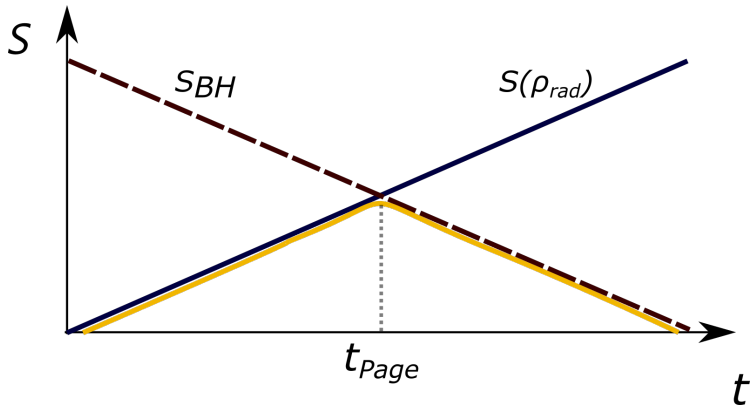


Figure 3: A schematic description of Page's paradox. The point at which the entanglement entropy grows larger than the Bekenstein-Hawking entropy is called the Page time, and the orange curve shows the predicted behaviour of $S(\rho_{rad})$ for a quantum thermodynamic system. Time is measured with respect to an inertial observer at asymptotic infinity.

the macroscopic observables is $N$-dimensional, then $S_{\text{th}} = \ln N$. Note that $S_{\text{th}} \geq S(\rho_{\text{bh}})$ for any particular black hole state. The question is then whether we have,

$$S_{BH} \stackrel{?}{=} S_{\text{th}}. \tag{5.9}$$

To address the question, consider a black hole in a pure state which has not yet emitted radiation. After some time, the black hole system will be in a mixed state due to the production of thermal radiation, but unitarity dictates that the combined state remains pure, giving $S(\rho_{\text{bh}}) = S(\rho_{rad})$. As we shall see in section 7.1, the $S(\rho_{rad})$ grows with time, while the Bekenstein-Hawking entropy shrinks due to energy loss through radiation. Thus, one inevitable reaches a contradiction where $S(\rho_{rad}) = S(\rho_{\text{bh}}) > S_{\text{th}}$, as illustrated in Fig. 3. The figure also shows the expected behaviour for a quantum thermodynamic system, as predicted by Page in [3]. Replicating this behaviour in a black hole model is the central topic of part II of the thesis.

# 6 Entropy in QFT and Gravity

In this section we discuss entanglement entropy in more detail. First, we show how a technique called the *replica trick* can be applied to compute the entanglement entropy of quantum fields. For conformal field theories (see appendix A) the replica trick gives an analytic result, which we apply in later sections. Then, we briefly introduce the AdS/CFT correspondence, which gives a relation between the entanglement entropy of certain conformal field theories and geometrical surfaces in negatively curved space. We show how this relation derives from the replica trick in a specific example. Including quantum corrections to the AdS/CFT result leads us to the generalized entropy formula eq. (6.29), which plays a central role in part II of the thesis.

For simplicity, we work in Euclidean time-signature in this section. However, the results generalize to the Lorentzian case [23, 24, 25].

## 6.1 The Replica Trick

To describe the state of a quantum field on some spacelike region $A$, one must first trace out the degrees freedom on the complementary region $\bar{A}$. The resulting state is described by a density matrix $\rho_A$, for which we wish to compute the entanglement entropy eq. (5.4). To this end, one can utilize the identity

$$S_A = -\mathrm{Tr}_A(\rho_A \log(\rho_A)) = -\frac{\partial}{\partial n}\left[\ln\left(\frac{\mathrm{Tr}\rho_A^n}{(\mathrm{Tr}\rho_A)^n}\right)\right]\Bigg|_{n=1}, \tag{6.1}$$

which is straightforward to show using standard rules of differentiation. For convenience, we introduce the label $Z_A(n) \equiv \mathrm{Tr}\,(\rho_A^n)$. As shown below, the replica trick is a method for evaluating $Z_A(n)$ by expressing it as a path integral over $n$ copies of the spacetime. One can then analytically continue the result to non integer $n$, and apply eq. (6.1) to compute the entanglement entropy[14]. As we shall see, analytic results can be obtained for conformal quantum field theories.

To demonstrate how the replica trick is applied, consider the example of a real scalar field $\Phi(t, x)$ on flat spacetime. For this case, the field operator defines a basis of eigenstates at $t = 0$ obeying $\Phi(x, 0)\,|\phi\rangle = \phi(x)\,|\phi\rangle$, where $\phi(x)$ is a real scalar function. The Euclidean time evolution of such a state is given by

$$|\phi(t)\rangle = e^{-tH}\,|\phi\rangle\,. \tag{6.2}$$

Just as in Lorentzian signature, we can write the transition amplitudes between two eigenstates as a path integral on the from

$$\langle\phi'(t')|\phi(0)\rangle = \langle\phi'|\,e^{-t'H}\,|\phi\rangle = \int_{(t=0):\,\phi(x)}^{(t=t'):\,\phi'(x)} D\phi\, e^{-S[\phi]}, \tag{6.3}$$

where the classical fields $\phi'(x)$ and $\phi(x)$ give the boundary conditions at $t = t'$ and $t = 0$ respectively. Note that by sending $t'$ to infinity, we obtain the transition amplitude between $|\phi\rangle$ and the vacuum, since the Euclidean time evolution suppresses all higher energy components of the state $|\phi'\rangle$ [27]. Thus, we can obtain a path integral expression for an element of the vacuum density matrix

$$\rho_{\phi_i\phi_j} = \langle\phi_i|0\rangle\langle 0|\phi_j\rangle = N\int_{\mathcal{C}(\phi_i,\phi_j)} D\phi\, e^{-S[\phi]}, \tag{6.4}$$

[14]Note that one should establish the uniqueness of the analytic continuation (see for example [26]), which is a topic that we will not discuss in this thesis.

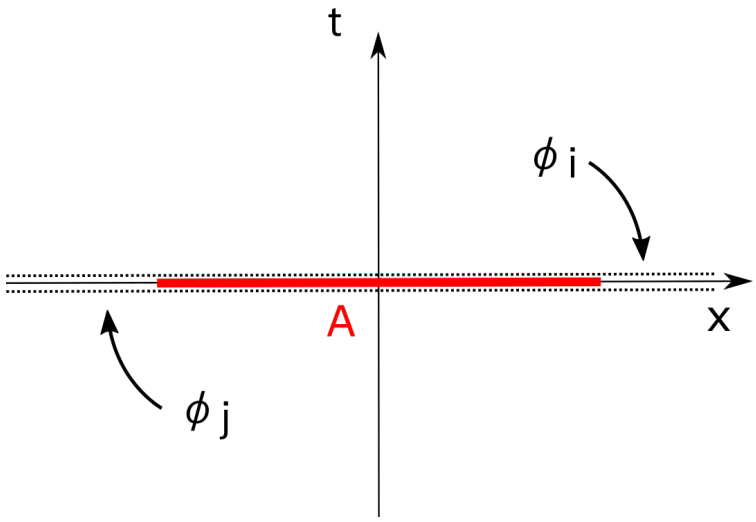


Figure 4: The boundary conditions of the path integral eq. (6.4), where $\phi_i$ and $\phi_j$ are the condition at $t = 0^+$ and $t = 0^-$ respectively.

where $\mathcal{C}(\phi_i, \phi_k)$ indicates that we integrate over the upper half-plane with boundary condition $\phi|_{t=0^+} = \phi_i$ and on the lower half-plane with $\phi|_{t=0^-} = \phi_j$ (see Fig. 4). The normalization factor $N$ in eq. (6.4) is not important, as it cancels out when we compute the entropy.

We can now restrict ourselves to the region $A$ by tracing out the complementary region $\bar{A}$. This is done by setting $\phi_i|_{\bar{A}} = \phi_j|_{\bar{A}}$ and integrating over all field configurations on $\bar{A}$ while keeping the fields fixed on $A$. This "glues" together the cut on $\bar{A}$ in the path integral, leaving

$$\rho_{A\,\phi_i\phi_j} \propto \int_{\mathcal{C}_A(\phi_i,\phi_k)} D\phi\, e^{-S[\phi]}, \tag{6.5}$$

where $\mathcal{C}_A(\phi_i, \phi_j)$ indicates that we impose the same boundary conditions as in eq. (6.4), but only on the interval $A$. Similarly to how we took the partial trace, the product $\rho_A^2$ is obtained by taking two copies of the cut path integral, with the boundary condition on the upper part of one cut equal to the lower part of the other, and integrating over all those field configurations. The result can be thought of as a single path integral on a manifold consisting of two copies of the original spacetime, glued along the cut on $A$. In fact, one can repeat this procedure for $n$ products of the density matrix, obtaining a manifold on the form shown in Fig. 5. The trace $Z_A(n) = \mathrm{Tr}(\rho_A^n)$ is then a obtained gluing together the two remaining boundaries, giving a path integral on the entire $n$-branched manifold. Thus, we have carried out the replica trick for a free scalar by expressing $Z_A(n)$ as a path integral on the form

$$Z_A(n) \propto \int_n D\phi\, e^{-S[\phi]}, \tag{6.6}$$

where we integrate over the $n$-copy manifold in Fig. 5.

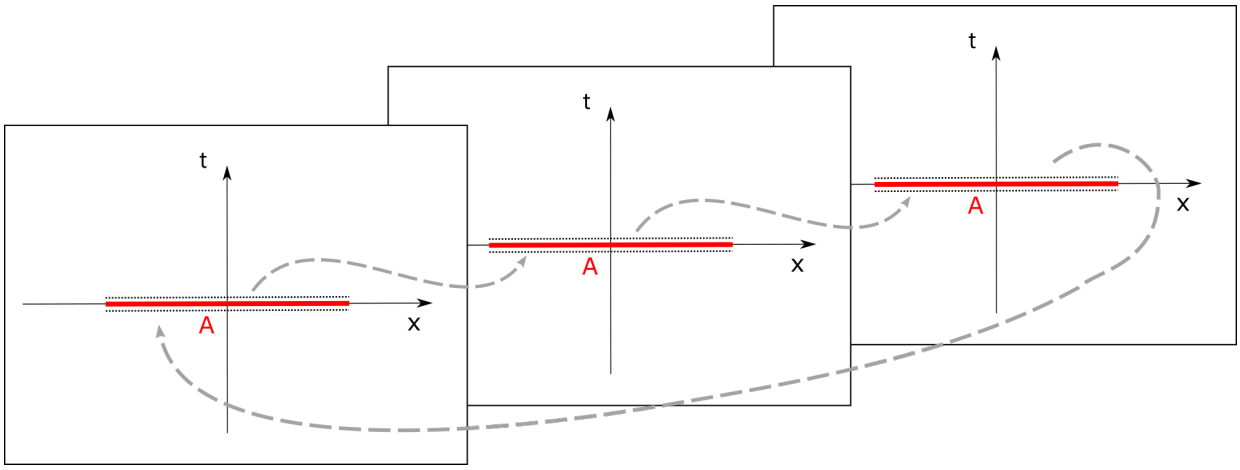


Figure 5: An illustration of the replica manifold $M^n$, obtained by sowing together $n$-copies of flat Euclidean space on the cut along $A$. The path-integral on the whole manifold gives $\mathrm{Tr}\rho_A^n$. Each copy corresponds to a power of the density matrix, and final gluing between the first and last copy corresponds to taking the trace. Here, we have set $n = 3$.

*Explicit result from conformal symmetry:*

The free massless scalar in 1+1 dimensions belongs to a class of special quantum field theories, called conformal field theories (see appendix A). For such theories, correlation functions of certain operators become very simple (see (A.10)). We can utilize this fact to obtain an explicit result from the replica trick for conformal fields. To this end, we apply a result from [28], which expresses eq. (6.6) as a two-point function of operators defined on the original Euclidean spacetime. The operators, $\mathcal{T}_n$ and $\bar{\mathcal{T}}_n$, are conformal primaries (see appendix A), meaning that their behavior under conformal transformations is given by eq. (A.12). The operators are inserted at the endpoints of the interval $A$, and have the same conformal weights

$$h_n = \tilde{h}_n = \frac{c}{24}(n - \frac{1}{n}), \tag{6.7}$$

where the constant $c$ is the central charge, which is one of the characteristics of a given conformal field theory (see appendix A) (note that $c = 1$ for the free massless scalar). These are known as twist-operators, and can be introduced in general when there is a global internal symmetry of the action, which in this case is the cyclic permutation symmetry between the fields on the different spacetime copies [28]. The main consequence of this result is that eq. (6.6) is determined by its behaviour under conformal transformations. Explicitly, we can introduce the Euclidean light cone coordinates $u = x+it$ and $v = -x+it$ (see appendix A) and label the endpoints of the interval as $(u, v)$ and $(u', v')$. We then see from eq. (A.14) that

$$Z_A(n) = \langle \mathcal{T}_n(u, v)\bar{\mathcal{T}}_n(u', v')\rangle \propto \left(\frac{|u - u'||v - v'|}{\epsilon}\right)^{-c(n-1/n)/6}, \tag{6.8}$$

where $\epsilon$ is a UV-cutoff [28]. As can be seen from eq. (6.9) below, changing the cutoff shifts the entropy by an overall constant, but does not alter the dependence on the interval length. One can now apply eq. (6.1) to eq. (6.8), obtaining

$$S_A = \frac{c}{6} \ln \left( \frac{|u-u'||v-v'|}{\epsilon} \right) + C', \tag{6.9}$$

where $C'$ is constant that depends on the proportionality constant in eq. (6.8). This result applies to any conformal field theory in 1+1 dimensions. The constant $C'$ can be absorbed in the UV-cutoff, so we neglect it in later sections.

Later, we will need to compute entanglement entropies in the static mirror setup from section 4. Therefore, we are interested in the case where the spacetime is the Euclidean half-space defined by $x > 0$. If the interval is on the form $A = [0, x)$, we can again use the results of [28], which in this case gives $Z_A(n)$ as the expectation value of a single primary operator inserted at the endpoint of the interval. Similarly to eq. (6.8), the scaling behaviour of the primary operator gives

$$S_A = \frac{c}{6} \ln \left( \frac{|u-v|}{\epsilon} \right) + C'', \tag{6.10}$$

where we will once again neglect the constant $C''$ in later sections.

## 6.2 Entropy in the AdS/CFT correspondence:

The AdS/CFT correspondence gives a connection between certain conformal field theories with large central charge and spacetimes geometries with negative cosmological constant (known as Anti-de Sitter, or AdS) [29, 30, 31]. A useful example is the vacuum of a holographic CFT[15] living on flat $d+1$-dimensional spacetime, which corresponds to an empty AdS universe with metric

$$ds^2 = \frac{l^2}{z^2} \left( dz^2 + dx_1^2 + \ldots + dx_d^2 + dt^2 \right), \tag{6.11}$$

where $l$ is a length-scale that fixes the cosmological constant and $z > 0$. A common feature among AdS spacetimes is the presence of an asymptotic boundary (in this case $z \to 0$) where one has to impose boundary conditions for fields in the interior. For a class of conformal field theories living on the asymptotic boundary of AdS, the correspondence then maps sources $J$ in the boundary theory to asymptotic boundary conditions for fields in the interior, giving a relation (schematically) on the form

$$Z_{CFT}[J] \approx e^{-I_{\text{grav}}[g_J]}, \tag{6.12}$$

[15] *Holographic* meaning simply that the CFT has a corresponding AdS description. A necessary condition is that the CFT has large central charge. More details about conditions for holographic CFT's can be found in [32].

where $Z_{CFT}$ is the generating functional of the conformal field theory and $I_{\text{grav}}$ is the standard gravitational action evaluated on the metric $g_J$ which depends on the boundary sources. The central charge of the boundary theory is related to the bulk geometry by

$$c = \frac{3l}{2G_N}. \tag{6.13}$$

In the above case of the vacuum on flat space, the energy scale of the boundary theory is closely linked to the $z$-dimension of (6.11), and we can impose the UV-cutoff of the previous section by setting $z > \epsilon$.

As a consequence of the correspondence between the generating functional of the boundary theory and the gravitational action in the interior, several field theoretic quantities can be more easily computed in the AdS spacetime. For our purpose the most important example is the entanglement entropy of a CFT on some boundary region $A$, given by

$$S_{\text{CFT}} \approx \underset{\Sigma}{\text{Min}} \left[ \underset{\Sigma}{\text{Ext}} \frac{A(\Sigma)}{4G_N} \right], \tag{6.14}$$

which is known as the Ryu-Takayanagi formula [33][16]. Here, $A(\Sigma)$ is the area of a co-dimension two surface[17] $\Sigma$, which is anchored on the boundary of the region $A$ (for an example with the metric eq. (6.11) and $d = 1$, see Fig. 6 ). The surface $\Sigma$ is chosen such that its area is extremized. If there are several surfaces satisfying eq. (6.14), the formula says to take the one that gives the smallest resulting area. Below we develop some intuition for this formula based on specific examples, and later we discuss its origin for a specific class of metrics (following along the lines of the more general proof in [34]).

### *Example: Vacuum, flat space*

Let us consider a conformal field theory on flat two-dimensional Euclidean space, as in the previous section. As mentioned above, the corresponding AdS metric is eq. (6.11) with $d = 1$. Here, we briefly confirm that the Ryu-Takayanagi formula reproduces the entropy result eq. (6.9) for an interval $A$.

In three dimensions a co-dimension two surface is just a curve, so the extremality condition in eq. (6.14) reduces to the problem of finding the locally shortest curve (i.e. the geodesic) that connects the endpoints of the interval. By time-symmetry of the metric, the curve must be located in the $t = 0$ plane. It is straightforward to show that the geodesic is a semi-circle with radius $a/2$, where $a$ is the length of the interval, so we get

$$z = \frac{a}{2}\sqrt{1 - \left(\frac{2x}{a}\right)^2}. \tag{6.15}$$

[16]The corresponding formula in Lorentzian signature is known as the Hubeny-Rangamani-Takayangi formula [23].

[17]For an $n$-dimensional surface in an $N$-dimensional spacetime, the co-dimension is simply defined as the difference $N - n$.

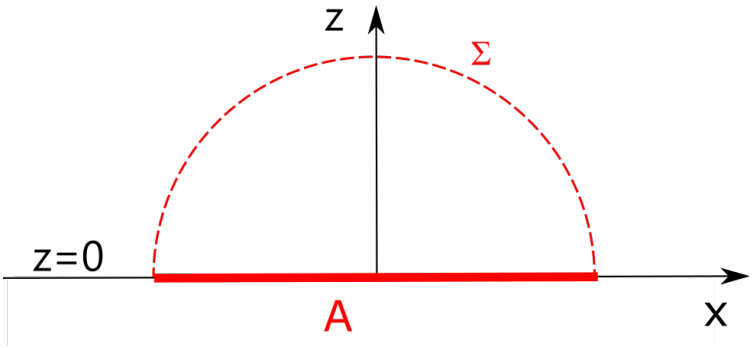


Figure 6: A constant time slice of vacuum AdS geometry eq. (6.11) at $t = 0$ with $d = 1$. The extremal surface that connects the endpoints of the interval $A$ on the boundary ($z = 0$) is the semi-circle $\Sigma$. The Ryu-Takayanagi formula eq. (6.14) agrees with the general CFT prediction eq. (6.9) for a single interval in the vacuum.

The length of the geodesic can then be evaluated as

$$S_A = \frac{A(\Sigma)}{4G_N} = \frac{1}{4G_N}\left[2\int_\epsilon^{\frac{a}{2}} \frac{l}{z}\sqrt{1+\left(\frac{dx}{dz}\right)^2}\,dz\right] = \frac{c}{3}\ln\left(\frac{a}{\epsilon}\right), \tag{6.16}$$

where the UV-cutoff $\epsilon$ appears in the calculation as the restriction $z > \epsilon$. Evidently, eq. (6.14) gives the correct entropy in this case. This result served as a non-trivial cross-check of eq. (6.14) in the original paper by Ryu and Takayanagi [33].

*Example: Black hole*

In this example, we build some intuition for extremal surfaces by studying a class of black hole metrics. It turns out that there is an extremal surface at the horizon, which plays a role in later discussions in part II of the thesis.

Consider the class of black hole metrics of the form

$$ds^2 = f(r)dt^2 + f^{-1}(r)dr^2 + h_{ij}(r,\theta)d\theta^i d\theta^j, \tag{6.17}$$

where $f(r)$ is a real function and $h_{ij}$ is the metric for the spacelike directions transverse to $r$. We assume the coordinates can be defined in such a way that $f(0) = 0$, meaning that the horizon $\Sigma$ is located at $r = 0$. We will show that the horizon is extremal. To start, we compute the trace of the extrinsic curvature

$$K \equiv \nabla_i \hat{n}^i, \tag{6.18}$$

where $\hat{n}$ is a unit vector field normal to $\Sigma$. The indices $i$ run over the tangent directions of the horizon. Intuitively, we can think of the trace of the extrinsic curvature as the divergence of $\hat{n}$ with respect to these tangent directions. If $K = 0$, the area will be left

unchanged by a perturbation along $\hat{n}$ [18]. Choosing the vector field $\hat{n}_r$ to lie along the $r$-direction we find

$$K = \Gamma^i_{ri}\sqrt{f(r)} = -\frac{\partial_r h}{2h}\sqrt{f(r)}, \tag{6.19}$$

where $\Gamma^i_{jk}$ denotes the usual Christoffel symbols. Assuming that $h_{ij}$ is non-degenerate at $r = 0$, we get $K = 0$, since $f(0) = 0$. Thus, the area of the horizon $\sigma$ is stationary under perturbations in the radial direction.

Note also that the metric eq. (6.17), a deformation of $\Sigma$ in the $t$ direction will not change the area, so the area must also be stationary under small perturbations in this direction. Since the $r$ and $t$ directions span the normal space of the horizon, the area is stationary under all normal perturbations, so the horizon is extremal.

## 6.3 AdS/CFT and the Replica Trick

To understand how the entropy formula eq. (6.14) follows from the replica trick, we derive it in the case of the black hole metric eq. (6.17) from the previous example. That is, we show how the replica trick applied to the right hand side of eq. (6.12) gives the area of the extremal surface at the horizon. For detailed proofs of eq. (6.14), we refer the reader to [34, 36].

In order to apply the replica trick to eq. (6.17), we first construct the gravitational version of the $n = 1$ replica manifold. To this end, we make the coordinate change $r = f'(0)\rho^2/2$ and $t = \left(f'(0)/\sqrt{2}\right)^{-1} t$, which puts the metric at $r \approx 0$ in the form

$$ds^2 \approx \rho^2 dt^2 + d\rho^2 + h_{ij}(r, \theta) d\theta^i d\theta^j. \tag{6.20}$$

Focusing on the $t$ and $\rho$ components we see that the metric is that of the plane in polar coordinates, with angular periodicity

$$t \sim t + 2\pi. \tag{6.21}$$

We assign this periodicity to the time coordinate, which corresponds to taking the trace in the replica trick. If one does not give the $t$-coordinate this specific periodicity, the geometry is that of a cone instead of a plane, and the curvature will have an unphysical singularity at $\rho = 0$[19].

We then apply the replica trick by extending the period of $t$ to $2\pi n$, which is analogous to the $n$-fold gluing in section 6.1. The metric on the replica-manifold differs from eq. (6.20)

[18] Intuitively, this is similar to how perturbing a closed surface along a divergence-free vector field leaves the volume intact, as seen e.g. in standard derivations of Liouville's theorem, as can be seen in many textbooks including [35].

[19] This is known as a conical singularity. It gives a divergent contribution to the scalar curvature that depends on the opening angle of the cone (cf. eq. (6.25)).

by a factor of $n^{-2}$ in front of $dt^2$, giving[20]

$$ds^2 \approx n^{-2}\rho^2 dt^2 + d\rho^2 + \dots . \tag{6.22}$$

We must then compute the action integral

$$I_{\text{grav}}[g] = \frac{1}{16\pi G_N}\int d^\mu x\sqrt{g}R, \tag{6.23}$$

on this $n$-copy replica-manifold. Using the symmetry in the $t$-direction we can write the action as $n$ times the action on a single copy

$$I_{\text{grav}}(n) = n\left[I_{\text{grav}}(n)\right]_{t\sim t+2\pi}, \tag{6.24}$$

where the right hand side differs from the $n = 1$ solution due to the $n$-dependence of the metric. Still, we can relate the right hand side to the $n = 1$ solution by studying the scalar curvature. Due to the mismatch between metric and the periodicity on the right hand side, the near horizon geometry is that of a cone with opening angle $2\pi/n$. The conical geometry has the same scalar curvature as the original manifold, except for at $r = 0$, where there is a delta-function contribution of the form

$$R_1 = R_n + 4\pi(1 - \frac{1}{n})\delta(r), \tag{6.25}$$

as shown in [37]. This corresponds to the sharp tip of the cone. We can then rewrite the integral over the $n$-branched manifold as an integral over the $n = 1$ manifold and an $n$-dependent contribution in the following way:

$$I_{\text{grav}}(n) = nI_{\text{grav}}(1) + \frac{(1-n)}{4G_N}\int_\Sigma d^2\theta\sqrt{h}, \tag{6.26}$$

where $h$ is the determinant of the transverse metric. The second term arises due to the delta function in eq. (6.25). This result also gives a straightforward analytic continuation to non-integer $n$. Inserting this into eq. (6.1) we find the entanglement entropy

$$S = \frac{A(\Sigma)}{4G_N}, \tag{6.27}$$

where $A(\Sigma)$ is the area of the horizon. The area is given by the integral in the rightmost term in eq. (6.26). We see that the area in eq. (6.14) arises because the replica trick affects the metric near the extremal surface at $r = 0$.

### *Quantum Correction:*

The area term in 6.14 gives the leading order contribution to the entanglement in orders of $\sim$ (in particular, a factor of $\sim^{-1}$ can be reintroduced by dimensional analysis). However,

[20] If the metric did not adapt to the change of periodicity, we would again get an unphysical conical singularity at $r = 0$.

there are sub-leading quantum contributions to the entanglement entropy. It is shown in [38] that the next-to-leading order contribution is given by the entanglement entropy of matter fields in AdS, giving a quantum correction on the form

$$S_{\text{correction}} = S_{\text{semi-classical}}(\Sigma), \tag{6.28}$$

where $S_{\text{semi-classical}}(\Sigma)$ is computed on the domain between $\Sigma$ and the asymptotic boundary (e.g. the black hole exterior if $\Sigma$ is the horizon). Note that $S_{\text{semi-classical}}$ is the entanglement entropy of matter fields in AdS, not the entropy of the CFT on the boundary. A priori it is not clear how to add this quantum contribution to eq. (6.14), as it could be placed either inside or outside the extremization. The latter case is argued in [39], giving the so-called *generalized entropy*

$$S_{\text{gen}} = \underset{\Sigma}{\text{Min}} \left( \underset{\Sigma}{\text{Ext}} \left[ \frac{A(\Sigma)}{4G_N} + S_{\text{semi-classical}}(\Sigma) \right] \right), \tag{6.29}$$

which is the case that we explore further in Part II of the thesis. In many cases, the quantum correction just slightly perturbs the position of an un-corrected extremal surface. However, the correction can also give rise to extremal surfaces that are not close to un-corrected surfaces. In the case of quantum fields interacting with gravity, eq. (6.29) can give a contribution to the entropy which is absent in naive semi-classical calculations, as we shall see in part II of the thesis. We refer to a surface $\Sigma$ that extremizes eq. (6.29) as a quantum extremal surface. If a quantum extremal surface is closed, we refer to its interior as a quantum extremal island.

# Part II

# Quantum Extremal Islands

## 7 Generalized Entropy

After the introduction of the quantum corrected entropy eq. (6.29), [5] and [6] applied the formula to models of evaporating black holes. They found that the generalized entropy reproduces all the important features of the Page curve. The result is surprising, as it has often been expected that one would need a full theory of quantum gravity to obtain the Page curve [22]. Instead, the generalized entropy is showing that just semi-classical calculations could be sufficient. Furthermore, the formula was applied in asymptotically flat spacetimes (see e.g. [40]), where no microscopic description (in the sense of AdS/CFT) is known, and still the Page curve was reproduced. This could indicate that eq. (6.29) applies for quantum fields coupled to gravity more generally, and not only for AdS/CFT [22]. In this section, we first give a qualitative description of how eq. (6.29) gives rise to the Page curve, followed by an explicit derivation in section 7.1 along the lines of [40].

The prescription for applying eq. (6.29) to a general spacetime region is as follows. First, define the the region of interest by introducing a boundary (illustrated as a red dotted curve in Fig. 7). Then choose a time slice[21] and search for quantum extremal surfaces in the interior of your region by extremizing the generalized entropy. If one is found which minimizes the generalized entropy, apply eq. (6.29) to the exterior of the surface (the blue region to the right in Fig. 7). If no quantum extremal surface is found, or if its generalized entropy exceeds that for the case without an island, apply eq. (6.29) to the whole region (the blue region to the left in Fig. 7). In the latter case, $\Sigma$ is called the empty surface.

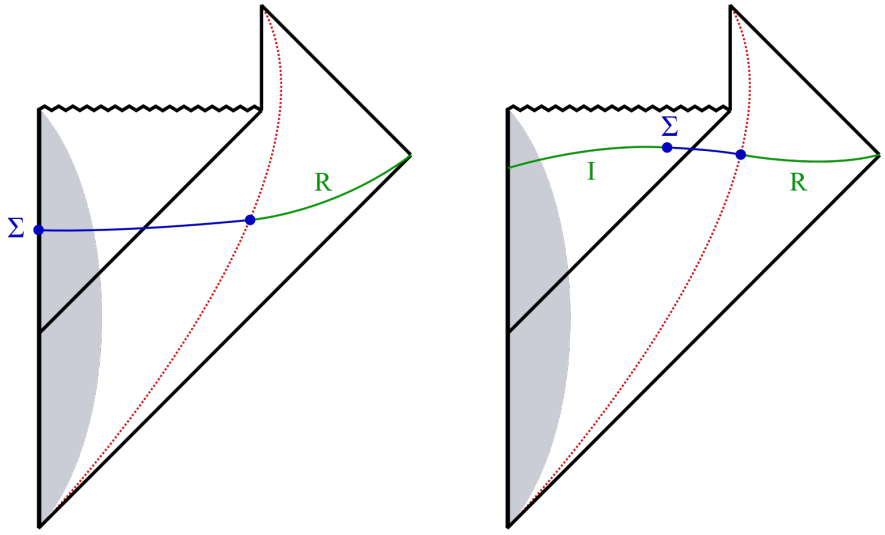


Figure 7: Two Penrose diagrams showing an evaporating black hole formed from collapse. The gray region represents collapsing matter, and the red curve is a boundary that partitions the spacetime into two regions. In the left diagram, we have a time slice where the empty surface gives the dominant contribution to the generalized entropy. To the right, we have a time slice containing a quantum extremal island $I$. $R$ denotes the region outside the red boundary.

The mechanism by which eq. (6.29) reproduces the Page curve can be explained intuitively in the following way. Consider a region containing a newly formed black hole (the left hand side of the red curve in Fig. 7, which has not emitted much radiation, and assume first that the horizon is a (non-quantum) extremal surface (as discussed in section 6.2). The entropy of the radiation in the black hole region will be small at early times, so we conclude that there will be a quantum extremal surface near the horizon soon after the black hole forms. Consider then the empty surface, which gives zero area contribution to the entropy. At early times this surface is the minimal quantum extremal surface, as the horizon area gives a much larger contribution to the entropy than the radiation. Consequently, at early times

[21] *Time slice* here meaning a spacelike surface of co-dimension one which extends from the origin of coordinates out to spacelike infinity. Two examples of time slices are illustrated Fig. 7, corresponding to the union of the blue and green regions. Intuitively, such a surface defines "an instant of time".

the entropy of our region will be given solely by the entropy of radiation (with a potential contribution from the collapsing matter), which increases as more quanta are emitted. This continues up to the point where the entropy of the radiation is comparable to the black hole area, at which point the near-horizon surface becomes dominant. This surface then gives a decreasing area-term which reproduces the decreasing part of the Page curve.

One might also wish to compute the generalized entropy for the radiation region $R$ outside the red dotted curve in Fig. 7. According to [22], this is given by the formula

$$S_{\text{rad}} = \operatorname*{Min}_{\Sigma} \left( \operatorname*{Ext}_{\Sigma} \left[ \frac{A(\Sigma)}{4G_N} + S_{\text{semi-classical}}\left(I \cup R\right) \right] \right), \tag{7.1}$$

where $I$ is the island in the black hole region. By definition, this quantity is equal to the generalized entropy. In the case that the empty surface dominates in eq. (7.1), the semi-classical entropy is computed on the region $R$. In the case that the island dominates, the semi-classical entropy is computed on the region $I \cup R$.

A few comments about eq. (7.1) are in order. Firstly it does not derive directly from the AdS/CFT-correspondence, as in this case there is no region $R$ outside the asymptotic boundary. Secondly, when there is a quantum extremal surface in the complement of R, the radiation inside the island is included in the semi-classical entropy calculation. This is surprising, as it implies that the region $R$ and the interior of the black hole is becoming less correlated (according to eq. (7.1)), while not supplying a physical mechanism by which these correlations disappear. Note that one can still obtain the Page curve for the black hole region using the generalized entropy, while remaining agnostic about the entanglement of the radiation region.

## 7.1 Islands in the RST model

In this section we show how generalized entropy reproduces the Page curve in the Russo–Susskind–Thorlacius (RST) model [41], as seen in [40]. In particular, we will use this model to investigate an evaporating two dimensional black hole formed from collapse. An important feature of the model is that it takes into account the back-reaction on the metric due to semi-classical effects, and still allows for simple analytic solutions to the equations of motion. This allows us to study a black hole that reacts to the presence of Hawking radiation by decreasing in area over time. We show that the evaporating black hole reproduces the Page curve. The RST model is closely connected to the moving mirror models in section 4. A detailed introduction to the model can be found in [42].

### 7.1.1 Review of the RST model

*The RST action:*

In two dimensions, the usual gravitational action (eq. (6.23)) vanishes for all variations of

the metric[22], and therefore does not give sensible equations of motion. Instead, one can construct a two dimensional gravitational action by introducing an additional field called a dilaton which couples to the scalar curvature[23]. A well known dilaton action is

$$I_{dilaton} = \frac{1}{2\pi}\int d^2x\sqrt{-g}\,\left(e^{-2\phi}\left(R + 4(\nabla\phi)^2 + 4\lambda^2\right)\right), \tag{7.2}$$

where $\phi$ is the dilaton field, $R$ is the scalar curvature and $\lambda$ is a constant measured in units of inverse length. The quantity $e^{-2\phi}$ plays the role of a varying gravitational coupling. One can then add matter to the theory in the form of $N$ free massless scalar fields, with a combined action on the form

$$I_{\text{matter}} = -\frac{1}{4\pi}\sum_{i=1}^{N}\int d^2x\sqrt{-g}\,(\nabla f_i)^2, \tag{7.3}$$

where each $f_i$ is a scalar field. The reason for choosing $N$ fields will be discussed shortly. One then finally obtains the RST action by adding an $N$-dependent term which makes the theory analytically solvable in the certain interesting cases, as we shall see below. The final term takes the form

$$I_N = \frac{1}{2\pi}\int d^2x\sqrt{-g}\,\frac{N}{24}R\,(\phi - \phi_0), \tag{7.4}$$

where $\phi_0 = \frac{1}{4}(1 - \ln\left(\frac{N}{48}\right))$, giving

$$I_{RST} = I_{dilaton} + I_{\text{matter}} + I_N. \tag{7.5}$$

This gives a classical two-dimensional gravitational theory coupled to matter. Importantly, $I_{dilaton}$ and $I_N$ both couple the dilaton to the scalar curvature, giving an overall gravitational coupling of

$$\bar{\Omega} \equiv e^{-2\phi} + \frac{N}{24}(\phi - \phi_0). \tag{7.6}$$

*Equations of motion:*

In two dimensions we can always find coordinates $x^+$ and $x^-$ that put the metric on the form $ds^2 = -e^{-2\rho}dx^+dx^-$, which simplifies the action considerably. The resulting equations of motion take a particularly simple form if we use the remaining freedom in the definition of $x^\pm$ to set $\rho = \phi + \frac{1}{2}\ln\left(\frac{N}{12}\right)$. In terms of the normalized gravitational coupling $\Omega \equiv \frac{12}{N}\bar{\Omega}$, we get

$$\partial_+\partial_-\Omega = -\lambda^2, \tag{7.7}$$

$$\partial_\pm^2\Omega = -T_{\pm\pm}, \tag{7.8}$$

[22]This is because $\sqrt{-g}R$ is a total derivative in two dimensions, so a local change in metric will not affect the integral over the whole spacetime manifold.

[23]One way to obtain a dilaton theory is to take a four-dimensional gravitational action and impose spherical symmetry on the metric. In that case, one can integrate over the angular degrees of freedom to obtain a two dimensional action. The new action includes a dilaton field which gives the area of the transverse sphere in the higher dimensional theory [43].

where

$$T_{\pm\pm} = \frac{6}{N}\sum_{i=1}^{N}\partial_\pm f_i \partial_\pm f_i \tag{7.9}$$

is the classical matter stress tensor multiplied by an $N$-dependent normalization constant. The reason for expressing the equations of motion in terms of the normalized stress tensor eq. (7.9) is discussed below.

*Boundary conditions:*

One can see from eq. (7.6) that $\Omega \geq 0$ for all real values of $\phi$, so in order to maintain the realness of the dilaton we must restrict ourselves to the region where this inequality is satisfied. Since $\Omega$ is a varying function of the coordinates, the boundary $\Omega = 0$ is a one dimensional curve in spacetime. The restriction is implemented in the RST model by imposing reflecting boundary conditions on the fields at $\Omega = 0$, i.e.

$$f_i|_{\Omega=0} = 0. \tag{7.10}$$

This is exactly the Dirichlet boundary condition that we put on the modes in the moving mirror models in section 4. That is, one lets $\Omega = 0$ define the trajectory of a moving mirror. As we see below, the mirror trajectory is influenced by the presence of matter fields.

*Back-reaction:*

We would like to study the RST-model in the semi-classical approximation, where the scalar fields are quantized while the dilaton and the metric are kept classical. One would also like to include the back-reaction of the quantum fields on the classical quantities. This can be done by replacing the classical matter stress tensor in the gravitational equations of motion with the expectation value,

$$\partial_\pm^2 \Omega = -\langle T_{\pm\pm}\rangle. \tag{7.11}$$

This approach for including quantum effects in the gravitational dynamics is commonly used in the limit where fluctuations in the stress tensor are small [20]. The reason for giving the stress tensor a $N$-dependent normalization constant is that fluctuations around the expectation value

$$\langle T_{\pm\pm}\rangle \propto \frac{1}{N}\sum_{i=1}^{N}\langle T^i_{\pm\pm}\rangle \tag{7.12}$$

are suppressed by taking the limit $N \to \infty$, see e.g. [42]. Still, depending on the choice of state and observer, there will be non-vanishing quantum contributions to the expectation value due to the normal ordering effects described by eq. (A.16). For details on back-reaction in the large $N$ limit, we refer the reader to [42].

*Vacuum solution:*

Let us now investigate how the mirror trajectory behaves in the flat space vacuum. That is, let us assume that we are in flat Minkowski space with metric $ds^2 = dudv$, and let us

also assume that we are in the vacuum state defined with respect to these coordinates, such that $\langle T_{\pm\pm}\rangle_{uv} = 0$. To solve the equations of motion we go to the $x$-coordinates, giving

$$\partial_{\pm}^2\Omega = \partial_{+}^2\phi + (\partial_{+}\phi)^2, \tag{7.13a}$$

$$\partial_{+}\partial_{-}\Omega = -\lambda^2, \tag{7.13b}$$

where the right hand side of the first equation arises from transformation law (A.16). It is straightforward to show that

$$\Omega = -\lambda^2 x^+ x^- - \frac{1}{4}\ln\left(-4\lambda^2 x^+ x^-\right), \tag{7.14}$$

solves the equations of motion. Inserting this into eq. (7.6) and using $\rho = \phi + \frac{1}{2}\ln\left(\frac{N}{12}\right)$, we solve for $\rho$ and find the metric

$$ds^2 = -\frac{\lambda^2}{x^+x^-}dx^+dx^-. \tag{7.15}$$

Comparing to the metric in $(u, v)$-coordinates, we see that eq. (7.15) can be obtained from the coordinate transformation

$$v = \lambda^{-1}\ln\left(2\lambda x^+\right), \quad u = -\lambda^{-1}\ln\left(-2\lambda x^-\right). \tag{7.16}$$

The $uv$-coordinates only cover the patch $x^+ > 0$, $x^- < 0$, but this is no problem as the region of interest (defined by $\Omega \geq 0$) is contained in this patch. Inserting eq. (7.16) into eq. (7.14) and setting $\Omega = 0$ , one finds the trajectory

$$v - u = 0, \tag{7.17}$$

showing that the mirror is static in these coordinates. Thus, for a vacuum state in flat space we get exactly the static mirror setup discussed in section 4.

### *Entanglement entropy:*

In order to obtain the Page curve we must compute entanglement entropies for the matter fields $f_i$. To this end, we can apply the static mirror entropy formula eq. (6.10). Note that by using the scale-invariance of the theory, the formula can be generalized to curved space. That is, since eq. (6.10) is given by the expectation value of a single primary operator, we know that a re-scaling of the metric $ds^2 = dudv \to ds^2 = e^{-2\rho}dudv$ gives a factor of $e^{-2\rho}$ in the entropy (c.f. A.15), giving

$$S(u,v) = \frac{1}{6}\ln\left(\frac{|u-v|}{\epsilon\, e^{-\rho(u,v)}}\right), \tag{7.18}$$

where one can interpret the $\rho$-dependent factor has having the role of keeping the cutoff $\epsilon$ at fixed proper length[24]. This formula will allow us to compute the entropy of a region containing a black hole when the empty surface gives the dominant contribution to eq. (6.29).

[24]Note that in eq. (6.10) $u$ and $v$ are complex, due to our discussion being in Euclidean signature. However, here we are working in Lorentzian signature, so $u$ and $v$ are real. The two cases are related by the transformation $t \to it$.

In the presence of an island we will need to compute the entropy of an interval with two end-points (see Fig. 7). The appropriate formula follows from the replica trick, as shown in [44], and takes the form

$$S(u,v,u',v') = \frac{1}{6}\ln\left(\frac{|u-v||u'-v'||v-v'||u-u'|}{\epsilon^2 e^{-\rho(u,v)}e^{-\rho(u',v')}|v-u'||u-v'|}\right), \tag{7.19}$$

where the primed and un-primed coordinates give the two endpoints of the interval. Note that eq. (7.19) contains an additional power of the cutoff length compared to eq. (7.18), which corresponds to the introduction of the additional endpoint.

*Coherent states:*

The entropy formulas eq. (7.19) and eq. (7.18) hold for the vacuum state in static mirror setup, but to study quantum extremal islands we will need to compute entropies for more general states. To this end, we can construct coherent states based on classical solutions to the equations of motion, which turn out to preserve the vacuum entropy formulas. A detailed introduction to such states can be found in [45].

To construct a coherent state, consider first a single scalar field and let $f_i$ be a classical solution to the equations of motion. Here, we assume that $f_i$ is a localized left-moving wave-packet, and we will only be concerned with the times before the packet reflects off the mirror. As such, we can treat the solution as a function solely of the coordinate $v$. We then define the coherent state as

$$|f_i\rangle = \hat{U}(f_i)\,|0\rangle_{uv}\,, \tag{7.20}$$

where

$$\hat{U}(f_i) =: \exp\left(\frac{i}{\pi}\int dv\,\partial_v f_i \hat{f}_i\right) : \tag{7.21}$$

is a unitarity operator that is normal ordered with respect to the vacuum defined by the $uv$-coordinates. The coherent state has a number of interesting properties, for example it is relatively straightforward to show that the expectation value of the field operator is given by the classical solution, $\langle\hat{f}_i\rangle = f_i$. Similarly, in $uv$-coordinates the expectation value of the stress tensor expectation is given by classical stress tensor of the solution $f_i$. Considering the entropy of the coherent state on some interval $A$, we can first factorize the unitary operator as

$$\hat{U}(f_i) = \hat{U}_A(f_i)\hat{U}_{\bar{A}}(f_i), \tag{7.22}$$

where the factor $\hat{U}_A|_A = \hat{U}|_A$, and is given by the identity otherwise (and similarly for the complement $\bar{A}$). We can then write the density matrix on $A$ as

$$\rho_A(f_i) = \mathrm{Tr}_{\bar{A}}\left(\hat{U}_A(f_i)\hat{U}_{\bar{A}}(f_i)\,|0\rangle\,\langle 0|\,\hat{U}_A^\dagger(f_i)\hat{U}_{\bar{A}}^\dagger(f_i)\right) = \hat{U}_A(f_i)\left(\mathrm{Tr}_{\bar{A}}\,|0\rangle\,\langle 0|\right)\hat{U}_A^\dagger(f_i), \tag{7.23}$$

where we have used the cyclicity of the trace. Inserting eq. (7.23) into eq. (5.4) and using the cyclicity of the trace once more, it is evident that the entanglement entropy of the

coherent state is equal to that of the vacuum[25]. That is, for a coherent state we can still utilize eq. (7.18) and eq. (7.19) (assuming that the mirror remains stationary, which will be the case below).

*Black hole:*

We can now investigate black holes in the RST-model. To this end, we can introduce infalling matter in a coherent state, whose stress tensor is that of a null shock wave, giving

$$\langle T_{++}\rangle_{uv} = M\delta(v), \qquad \langle T_{--}\rangle_{uv} = 0, \tag{7.24}$$

where $M$ is the total energy of the in-going matter [42]. It is important that the matter is coherent, as it allows us to use the entropy formulas eq. (7.18) and eq. (7.19) before the shock wave hits the mirror. Note that the delta function in eq. (7.24) just appears as an additional term in the vacuum-equation of motion, which can be accounted for by adding a corresponding step-function term to the vacuum solution, giving a black hole solution on the form

$$\Omega = -\lambda^2 x^+ x^- - \frac{1}{4}\ln\left(-4\lambda^2 x^+ x^-\right) - \lambda^{-1}M(\lambda x^+ - 1)\Theta(x^+ - 1), \tag{7.25}$$

where we have used the identity $\partial^2\left(x\Theta(x)\right) = \delta(x)$. The resulting mirror trajectory (defined by $\Omega = 0$) is sketched in Fig. 8. It will evidently remain stationary in the $uv$-coordinates (as in the vacuum case) for $x^+ < 1$, i.e. before the shock wave hits the mirror. At this point, the trajectory becomes spacelike, and remains so until

$$x^+_{evap} = \lambda M^{-1}(e^{4\lambda^{-1}M} - 1), \tag{7.26}$$

where it once again becomes timelike. The spacelike part of the mirror trajectory acts a black hole "singularity", where one can no longer impose sensible boundary conditions on the fields [40]. The $x^-$ coordinate of the point where the mirror trajectory becomes timelike defines the event horizon of the black hole, inside of which all non-spacelike trajectories will hit the singularity. We find

$$x^-_{horizon} = -\lambda^{-1}Me^{4\lambda^{-1}M}(e^{4\lambda^{-1}M} - 1)^{-1}. \tag{7.27}$$

When the horizon meets the mirror trajectory at $x^+_{evap}$, we say that the black hole has evaporated away. We will only be concerned with the regions outside the future light cone of this event.

*Hawking Radiation:*

We can also confirm that this black hole gives rise to outgoing Hawking radiation. First we note that in the limit $x^+ \to \infty$, the metric takes the form

$$ds^2 \approx \lambda x^+(\lambda x^- + \lambda^{-1}M)dx^+dx^-, \tag{7.28}$$

[25]This argument ignores certain subtleties regarding the UV-cutoff, which are discussed in [45]

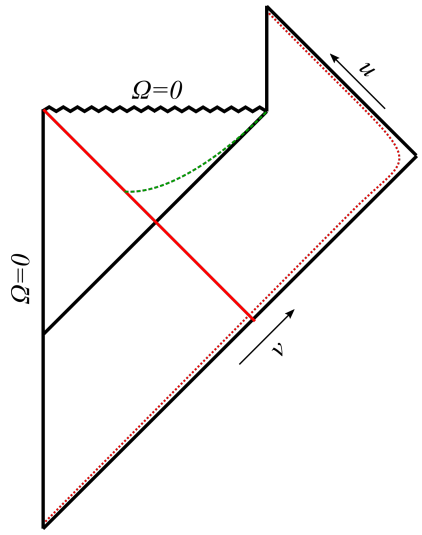


Figure 8: A Penrose diagram of an evaporating black hole in the RST-model. The red line represents the in-going null shock wave, and the green curve shows the trajectory of the island. The red dotted curve is the boundary of region for which we compute the generalized entropy, and is taken to be nearly lightlike at large $v$.

as follows from eq. (7.6) and the identity $\rho = \phi + \frac{1}{2}\ln\left(\frac{N}{12}\right)$. We can then transform to inertial coordinates $\bar{u}$ and $\bar{v}$ such that the above metric takes form $ds^2 = d\bar{u}d\bar{v}$. In this case, we get

$$\bar{v} = \lambda^{-1}\ln\left(\lambda x^+\right), \tag{7.29a}$$

$$\bar{u} = -\lambda^{-1}\ln\left(-\lambda x^- - \lambda^{-1}M\right), \tag{7.29b}$$

These are the coordinates of an inertial observer asymptotically far away from the black hole at late times. Applying eq. (A.16) to compute the stress tensor in the inertial coordinates, we find an out-going energy flux

$$\langle T_{--}\rangle_{\bar{u}\bar{v}} = \lambda^2\frac{\lambda^{-1}Me^{\lambda\bar{u}}(2+\lambda^{-1}Me^{\lambda\bar{u}})}{48\pi(1+\lambda^{-1}Me^{\lambda\bar{u}})^2}, \tag{7.30}$$

which is the energy of Hawking radiation measured by an inertial observer at asymptotic infinity. To leading order in $M$, the total out-going energy flux (obtained by integrating the energy flux up to the evaporation endpoint) is equal to the total black hole mass $M$. Note also that for $M \to \infty$ we have $\langle T_{--}\rangle_{\bar{u}\bar{v}} \approx \frac{\lambda^2}{48\pi}$, which can be compared with the thermal flux of eq. (4.13). We see that $\lambda$ controls the temperature of the outgoing Hawking radiation.

### 7.1.2 The Page curve

To obtain the Page curve, we must first express the generalized entropy in the RST-model. One difficulty is that in two-dimensions, a quantum extremal "surface" is only a single

point, and does not have an area. However, if we look at eq. (6.26) in two dimensions, we see that the two-dimensional "area term" takes the form $1/4G_N$. Accounting for the different normalization of the RST-action compared to eq. (6.23) and setting eq. (7.6) as the gravitational coupling (taking the place of $G_N^{-1}$), we get

$$\frac{A}{4G_N} \to 2\bar{\Omega}, \tag{7.31}$$

for the area term in the RST model[26]. Thus, for an interval with endpoints $P$ and $P'$ the generalized entropy takes the form

$$S_{island} = 2\Omega(P) + S(P, P'), \tag{7.32}$$

where the last term is given by eq. (7.19). We can now search for quantum extremal islands by extremizing the generalized entropy. To this end, we take the outer endpoint $P'$ to be anchored on the boundary of the black hole region, while we vary the other endpoint to locate the quantum extremal surface. We assume that the anchor point is placed at large $v'$ and that the island starts of close to the horizon. The resulting trajectory remains close to the horizon, so the approximation is justified. Solving $\partial_u S_{island} = 0$ and $\partial_v S_{island} = 0$ then gives us $v$ and $u$ in terms of the asymptotic coordinate $u'$,

$$v = -\ln\left[4\lambda^{-1}M - e^{-\lambda u}\right], \tag{7.33}$$

$$u = -\ln\left[-\lambda^{-1}MW_{-1}\left(-e^{-\lambda u'-4}\right)\right], \tag{7.34}$$

where $W_{-1}(x)$ is a branch of the Lambert W-function, defined by $W_k(xe^x) = x$. We must choose $k = -1$ since it is the only branch that is well defined for negative inputs. Thus, we have the trajectory of a quantum extremal island, shown in Fig. 8.

We must then compare the generalized entropy of the extremal island to the case without an island. In this case, the entropy of the black hole region is given by eq. (7.18). In the large $v'$ limit, we get

$$S_{no-island}(u', v') = \frac{1}{2}\ln\left(1 + \lambda^{-1}Me^{\lambda u'}\right) + \ln\left(\frac{v'-u'}{\epsilon}\right), \tag{7.35}$$

which now gives us all the ingredients needed to obtain the Page-curve. In particular, we can now insert the trajectory of the quantum extremal surface into eq. (7.32) and eq. (7.35) and determine the generalized entropy, given by

$$S_{gen} = \min\{S_{no-island}, S_{island}\}. \tag{7.36}$$

The result is plotted in Fig. 9 with respect to the asymptotically inertial coordinate $\bar{u}$ (defined above), assuming that the anchor point is on a nearly light-like trajectory such

[26]Intuitively, one could also interpret the coupling $\bar{\Omega}$ as an area in a higher dimensional theory, as is the case for the coupling $e^{-2\phi}$ in certain dilaton theories (as mentioned above). [43]

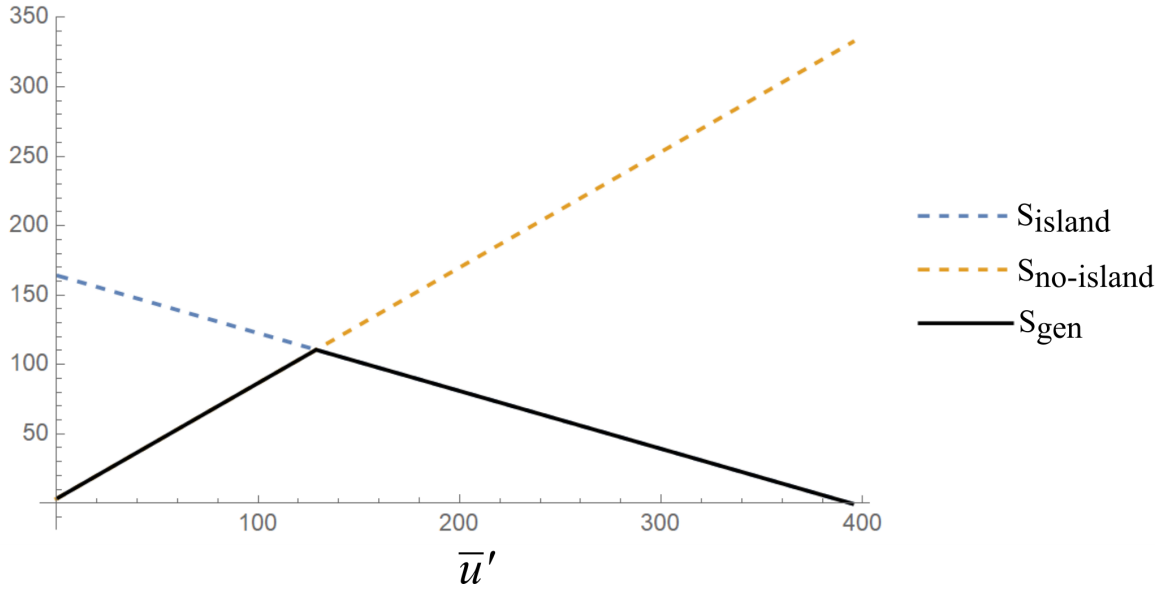


Figure 9: The Page curve for an evaporating black hole in the RST model. The blue and orange curves show the entropy of the island and no-island contributions, respectively. The result is plotted with respect to the affine time $\bar{u}'$ of an inertial observer asymptotically far away from the black hole. Here, we have set $N = 10$, $M = 100$ and $\epsilon = 1$.

that $\bar{v}'$ is approximately constant (see figure 8). The plot shows the expected Page-curve behaviour, with the Page time located at $\bar{u}' \approx 4M\lambda^{-2}/3$ for large $M$. One can note that the island entropy reaches zero at $\bar{u}' \approx 4\lambda^{-2}M - \ln\lambda^{-1}\,(\lambda^{-1}M)$, which is the expected evaporation endpoint for large $M$. The semi-classical approximation is likely to break down near the evaporation endpoint, but results at earlier times appear to be robust [42].

*Note on the UV-divergence:*

In the Fig. 9 we have set the UV-cutoff to $\epsilon = 1$, which ignores a subtle technical issue. In fact one has to be careful because the island entropy $S_{island}$ has a stronger dependence on the cutoff than $S_{no-island}$. In [40], as well as other well known derivations in dilaton theories (such as [5]), this issue is circumvented by introducing a corresponding UV-divergence through a renormalization of the dilaton, which cancels one power of $\epsilon$ in eq. (7.32), and thus gives the two entropies identical UV-behaviour. However, this procedure is not explained in depth in the aforementioned references, and is something that we would like to understand more thoroughly.

# 8 A Puzzle Regarding Islands

In the original formulation of the information paradox the matter in the interior of the black hole does not play a central role, as the interior degrees of freedom are traced out.

On the other hand, eq. (6.29) and eq. (7.1) make explicit use of semi-classical calculations in the interior. In particular, before the island contribution becomes dominant, eq. (6.29) states that the generalized entropy of spacetime region containing a black hole is given by the semi-classical entanglement entropy the entanglement entropy of the matter contained within the region. In this section we discuss this feature of the generalized entropy, and show explicitly that semi-classical calculations in the interior conflicts with the thermodynamic interpretation of the Bekenstein-Hawking entropy in the RST-model even before the Page time.

Consider first a thought-experiment where a black hole radiates up to a point right before the Page time. One could imagine gathering the emitted radiation, which we know is in a high entanglement state, and using it to form a new black hole. It would then seem that the emission of Hawking radiation from the new black hole could lead the total entanglement entropy of the matter in the interior to exceed the Bekenstein-Hawking entropy before the Page time (for the new black hole). This is particularly clear for large non-rotating black holes, for which the Page time is expected to occur near the half-way point of evaporation [4]. In that case, the entanglement entropy of the radiation that forms the new black hole would be nearly equal to its Bekenstein-Hawking radiation, and emission of Hawking radiation would allow the entanglement entropy of the matter in the interior to exceed the Bekenstein-Hawking entropy after a small amount of time. Since the entanglement entropy sets the lower bound for the thermodynamic entropy (see eq. (5.9)), we conclude that the thermodynamic entropy of matter in the interior can exceed the Bekenstein-Hawking entropy of the black hole.

A simple way to avoid this issue is to argue that the semi-classical states do not accurately describe the interior, and that a more fundamental theory of quantum gravity is needed to obtain behavior consistent with the Bekenstein-Hawking entropy. Indeed, that is a central reason for the interest in the generalized entropy, as it originates from a microscopic theory of quantum gravity (in the sense of AdS/CFT) and gives result consistent with the thermodynamic interpretation of the Bekenstein-Hawking entropy for a large variety of black holes (that is, by reproducing the Page curve). The puzzle is then why the generalized entropy still makes reference to semi-classical field states in the interior. To the best of our knowledge, no clear answer has emerged in the literature. An attempt to circumvent the question could be to avoid time-slices that penetrate deeply into the black hole interior. This would still allow one compute the generalized entropy of a black hole region, provided that time-slice can extend to the past of the formation of the horizon up until the appearance of the island.

## 8.1 Violating the Bekenstein-Hawking Bound in the RST Model

We can use the RST model to show an explicit example where the entanglement entropy of matter fields in the interior exceed the Bekenstein-Hawking entropy of the black hole. In particular, we can form a black hole from non-coherent matter, which modifies the entropy

formulas eq. (7.18) and eq. (7.19). As a result we find that the entanglement entropy in the interior can exceed the Bekenstein-Hawking entropy before the Page time.

Consider first the vacuum state defined with respect to the null coordinates $(U, V)$. We will construct a change of coordinates such that the shock wave is obtained from the stress tensor transformation law eq. (A.16). To this end, one makes use of the fact that

$$f(x) = \frac{Ax + B}{Cx + D} \implies \{f(x), x\} = 0, \tag{8.1}$$

where $A, B, C, D$ are constants and $f$ is known as a fractional linear transformation. We can then patch together two coordinate transformations of this form to obtain a stress tensor that is zero everywhere except at the patch. A version of this trick is utilized in a different context in [5]. We define $U(u) = u + w_0$ and

$$V(v) = \begin{cases} \dfrac{(w_0 + w_1)v + w_0 w_1}{w_1 + v}, & v > 0 \\ \dfrac{w_0^2}{w_0 - v}, & v \leq 0 \end{cases} \tag{8.2}$$

where $w_0$ and $w_1$ are positive constants. The resulting coordinate transformation has continuous first derivative, but a discontinuous second derivative which gives rise to a delta function in the Schwarzian. In particular, we have

$$\langle T_{--} \rangle_{uv} = M\delta(v), \tag{8.3}$$

where

$$M = \frac{w_0 + w_1}{w_0 w_1}. \tag{8.4}$$

Evidently, we are able to vary the coordinate transformation while keeping the energy of the shock wave fixed. This allows us to vary the entanglement entropy of the shock wave without modifying the behaviour of the black hole. In particular, transforming eq. (7.18) to the $uv$-coordinates we get

$$S(u, v) = \frac{N}{6}\left[\rho(u, v) + \ln\left(\frac{|U(u) - V(v)|}{\epsilon\sqrt{U'(u)V'(v)}}\right)\right], \tag{8.5}$$

where the derivative terms arise from the coordinate transformation[27]. We can now set the point $(u, v)$ on the horizon, and compute the entropy of the black hole interior[28]. To this end, note first that in the limit $w_1 \to \infty$ the coordinate transformation is simply a

[27] As is straightforward to show using eq. (6.9) and the transformation-law eq. (A.12).

[28] In [42], the interior of the black hole is defined as the region inside the "apparent horizon", defined by $\partial_+\Omega = 0$, instead of the event horizon. However, the two horizons are close near the Page time for large $M$, so this distinction is not needed to arrive at the result eq. (8.8). For a review of apparent horizons, see for example [46].

linear shift of the $v$ coordinate for $v > 0$ (i.e. after the shock wave), which leaves eq. (8.5) identical to eq. (7.19). In terms of $u$ and $v$, we get

$$S(u,v) \approx \frac{N}{6}\left[\rho(u,v) + \ln\left(\frac{v+\ln(M)}{\epsilon}\right)\right], \tag{8.6}$$

where we also assume that the shock wave energy is large. Similarly, taking the lower limit $w_1 \to M^{-1}$ gives

$$S(u,v) \approx \frac{N}{6}\left[\rho(u,v) + \ln\left(\frac{\ln(M)(1+Mv)}{\epsilon}\right)\right]. \tag{8.7}$$

Near the Page time quantum extremal surface lies approximately at the the horizon, where the Bekenstein-Hawking entropy takes the value $S_{BH} \approx \frac{2}{3}M$. We also get the corresponding value $v \approx 4M/3$, giving the difference between eq. (8.6) and eq. (8.7) as

$$\Delta S \approx M\ln(M). \tag{8.8}$$

Thus, near the Page time the entropy given by eq. (8.7) will exceed the Bekenstein-Hawking entropy by approximately a factor of $\frac{3}{2}\ln(M)$. Since the energy of the shock wave was kept fixed, this result gives a lower bound for the thermodynamic entropy of the matter in the interior even when the shock wave is in a coherent state. Thus, the thermodynamic entropy in the interior of a black hole formed from a coherent state can grow larger than the Bekenstein-Hawking entropy before the Page time.

# Part III

# Moving Mirrors and Quasiparticles

## 9 The Alba-Calabrese Formula

In this section, we introduce an approximation that has been used to compute entanglement entropy in quantum lattice systems, known as the Alba-Calabrese (AC) formula [47]. Essentially, the formula shows that entanglement entropy in certain cases behaves as if pairs of entangled *quasi-particles* propagate ballistically through the system. In the following section, we show results indicating that the AC formula can accurately reproduce the entropy of single and double intervals in a moving mirror model. To the best of our knowledge, the AC approach has not previously been applied in this setting. A detailed introduction to the AC formula can be found in [48].

Suppose we want to compute the entropy of an interval $A$ of length $a$, embedded in an infinitely long one-dimensional quantum system[29]. In particular, suppose that the system

[29] A simple example of such a system could be an infinite lattice of coupled harmonic oscillators, for

is in the ground state up to $t = 0$, where a sudden shock to the system occurs in the form of a rapid change in the Hamiltonian[30]. For a large variety of systems, the subsequent entanglement dynamics are accurately described by the quasiparticle picture developed in [47]. That is, the entanglement entropy is accurately described by a model where independent pairs of entangled quasiparticles are created uniformly throughout the system and propagate ballistically in opposite directions. The entanglement entropy of the interval $A$ is then approximately given by adding up the entanglement entropy of every quasiparticle on $A$. Since each quasiparticle is only entangled with its partner, a pair that is wholly inside or wholly outside the interval does not contribute to the entanglement of $A$. However, at a given instance of time there will be pairs where only one quasiparticle is located in the interior, giving a non-zero entanglement contribution.

For simplicity, assume we have a single kind of quasiparticle travelling at velocity $v_i$. The entanglement contribution from a single pair of quasiparticles is denoted $s_i$. Initially, at $t = 0$ there will be no quasiparticle pairs that contribute to the entanglement, as no quasiparticles have crossed the boundary of the interval. Then, for $t > 0$ we get a linear growth in entropy as quasiparticles cross the interval boundaries at a rate of $2v_i t$. This continues until the first quasiparticle pair completely leaves the interval at $t = \frac{a}{2v_n}$. At this point, the entanglement entropy stabilizes due to the balance of quasiparticles entering and leaving the interval. This behaviour is captured by the AC formula, which in our simple example takes the form [49]

$$S_A^i(t) = \begin{cases} 0 & \text{if} \quad t \leq 0, \\ 2s_i v_i t & \text{if} \quad 0 < t < \frac{a}{2v_i}, \\ s_i a & \text{if} \quad \frac{a}{2v_i} \leq t. \end{cases} \tag{9.1}$$

In general cases there will be several kinds of quasiparticles with different velocities $v_i$, and one would have to add up the contributions from each kind of quasiparticle. This would give a less trivial time dependence than the piecewise linear function above. However, the above expression is sufficient for our purposes.

# 10 Quasiparticle Picture for Moving Mirrors

For any CFT in a moving mirror model, we have a general expression for the entanglement entropy on a region with a single endpoint (see eq. (6.10)). On the other hand, results for more complicated regions generally depend on the specific theory under consideration. Therefore, it is interesting to check whether the quasiparticle picture can provide us with a general approximation for complicated domains. The quasiparticle approach to moving

which the Hamiltonian takes the familiar form $H = \frac{1}{2}\sum_n \pi_n^2 + m\phi_n^2 + (\phi_{n+1} - \phi_n)^2$, were $\phi_n$ and $\pi_n$ are the position and momentum operators of the $n$'th oscillator and $m$ is a mass parameter [48].

[30]For the harmonic lattice, this could be a sudden change in the mass paramater.

mirror models was explored recently in [50], and here we build on this work by constructing appropriate Alba-Calabreses formulas and testing them against analytic results for free massless fermions. We find that the formulas accurately reproduce the entanglement entropy of free fermions on a single and a double interval in the limit of high temperature radiation. Our AC formulas are derived without reference to a specific CFT, so one could hope that the results are general.

Recall from section 4 that a moving mirror trajectory is characterized by a function $p(u)$. Here, we consider a specific trajectory used in [50], where

$$p(u) = -\kappa^{-1} \ln \left(1 + e^{-\kappa u}\right). \tag{10.2}$$

This trajectory has the useful property that it becomes stationary at asymptotically early times, while producing a constant flux of thermal radiation at late times. In particular, for large $u$ we have $\langle T_{uu} \rangle \approx \frac{c\kappa^2}{48\pi}$, which can be compared with the thermal flux of eq. (4.13). Note that in the high-temperature limit $\kappa \to \infty$, the trajectory is close to stationary until $t = 0$, at which point it quickly becomes approximately linear.

Let us then compute the entanglement entropy for domain $A$ with a single endpoint $(u, v)$. By the same argument given around eq. (8.5) the appropriate formula takes the form

$$S_A(u, v) = \frac{c}{6} \ln \frac{|v - p(u)|}{\epsilon \sqrt{p'(u)}}, \tag{10.3}$$

which holds for a general CFT. We consider a static endpoint of the form $(u(t), v(t)) = (t - x_0, t + x_0)$, where $x_0$ is constant. Inserting this into eq. (10.3) and taking the high temperature limit $\kappa \to \infty$, we find (ignoring the contribution from the UV-cutoff $\epsilon$ for now)

$$S_A(t) \approx \begin{cases} 0 & \text{if} \quad t \leq x_0, \\ \frac{c\kappa}{12}(t - x_0) & \text{if} \quad x_0 < t, \end{cases} \tag{10.4}$$

which is highly suggestive of a quasiparticle interpretation along the lines of eq. (9.1). It seems as if a right-moving flux of quasiparticles is created near the mirror and travels out at the speed of light before hitting the interval, giving sudden linear increase in the entropy. This is argued more thoroughly in [50], where the authors show that quasiparticle pairs are created along the curve $p(u) + v = 0$ (i.e. the reflection of the mirror across the $u$ diagonal). In the large temperature limit this curve is equal to the mirror trajectory itself for $t > 0$ . We see from eq. (10.4) the each quasiparticle pair gives a contribution to the entanglement that only depends on the central charge and the temperature, namely $\frac{c\kappa}{12}$.

Using the results of the previous paragraph (which were shown in [50]) we can now construct Alba-Calabrese formulas for more general intervals. The reasoning is almost exactly the same as the one given for eq. (9.1), the only difference being that we consider a flux of right-moving quasiparticles originating near the mirror. It is a trivial exercise to show that

the appropriate formula for an interval $B = [b_1, b_2]$ is

$$S_B(t) \approx \begin{cases} 0 & \text{if} \quad t \leq b_1, \\ \frac{c\kappa}{12}(t - b_1) & \text{if} \quad b_1 < t \leq b_2, \\ \frac{c\kappa}{12}(b_2 - b_1) & \text{if} \quad b_2 < t. \end{cases} \tag{10.5}$$

As in eq. (10.4), the Alba-Calabrese result is zero until the first quasiparticle hits the interval. Then, there is linear growth until the first quasiparticle hits the second boundary of the interval. At this point the entanglement entropy stabilizes due to the balance of quasiparticles entering and leaving the interval. The case of region consisting of two disjoint intervals on the form $C = [c_1, c_2] \cup [c_3, c_4]$ follows along exactly the same lines.

The central question is then whether the Alba-Calabrese result gives an accurate description of the entropy for the regions with more than one endpoint. Analytic results do not exist for general CFTs, but we can test the above formulas using the solvable example of free massless fermions. The resulting equations for $S_B$ and $S_C$ can be found in [51]. For the interval $B = [b_1, b_2]$ we get,

$$S_B^{ferm}(t) = \frac{1}{12}\left[\ln\left(\frac{|b_1 - b_2|}{\epsilon^2}\right) + \ln\left(\frac{|p(t - b_1) - p(t - b_2)|}{\sqrt{p'(t - b_1)p'(t - b_2)}}\right) + \ln W(b_1, b_2, t)\right], \tag{10.6}$$

where

$$W(b_1, b_2, t) \equiv \frac{|(t + b_2) - p(t - b_2)||(t + b_1 - p(t - b_2)|}{|(t + b_1) - (t - b_2)||(t + b_2) - p(t - b_1)|}. \tag{10.7}$$

The analytic result for $C$ is more complicated and we do not give it explicitly here in order to avoid excessive clutter. The formula is readily available in [51]. The analytic results are compared with the Alba-Calabrese formula in Fig. 10 and Fig. 11. We see that the results are accurate at high temperatures. The fact that we need to go to high temperatures might not be surprising, as we know from quantum many-body systems that high temperatures give low correlation lengths, and we assumed from the beginning that quasiparticles originating at different positions are uncorrelated [48].

A final interesting question is then whether the Alba-Calabrese results are applicable to the RST-model. We will not be able to fully explore this question in this thesis, but we can show an encouraging initial results along the lines of eq. (10.4). In particular, consider the entropy formula (7.18) transformed to the coordinates $\bar{v}(v) = v$ and $\bar{u}(u) = \lambda^{-1}\ln\big(e^{-\lambda u} + \lambda^{-1}M\big)$, which give a flat metric in the limit $v \to \infty$ (see discussion around eq. (7.30)). The resulting entropy is

$$S(\bar{u}, \bar{v}) = \frac{N}{6}\left[\rho(u(\bar{u}), v(\bar{v})) + \ln\frac{|u(\bar{u}) - v(\bar{v})|}{\epsilon} + \frac{1}{2}\ln\left(1 + \lambda^{-1}Me^{\lambda\bar{u}}\right)\right]. \tag{10.8}$$

We can then repeat the above procedure where we fixed the endpoint as $(\bar{u}(t), \bar{v}(t)) = (t - x_0, t + x_0)$ for a constant $x_0$. The constant is chosen such that the endpoint stays outside the black hole. For late times $t$, we then find

$$S(\bar{u}(t), \bar{v}(t)) \approx \frac{N\lambda}{12}t. \tag{10.9}$$

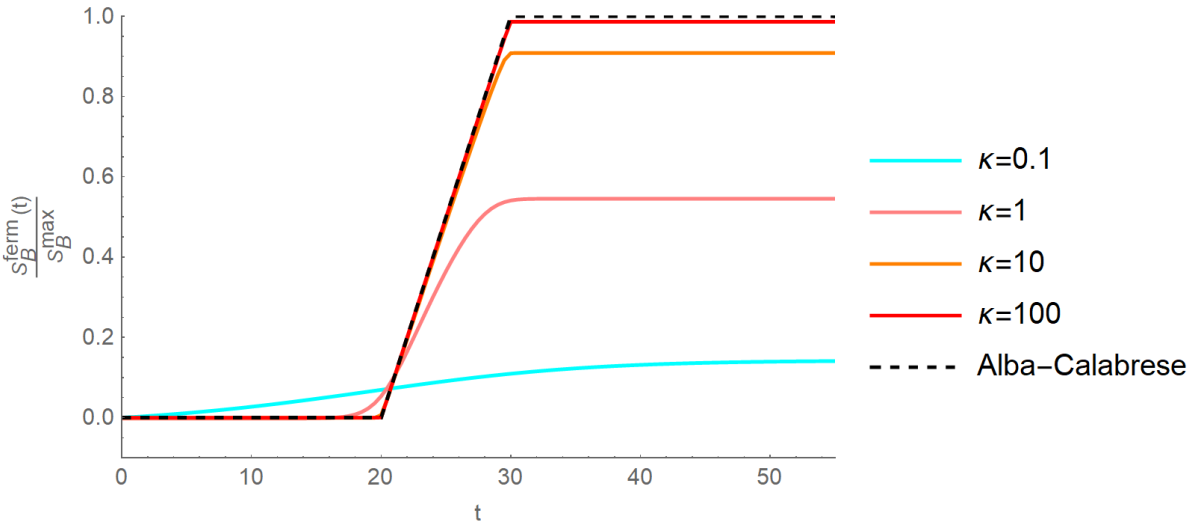


Figure 10: The entanglement entropy for a free massless fermion field on the interval $B = [20, 30]$ for different values of the temperature parameter $\kappa$, compared to the Alba-Calabrese formula (eq. (10.5)). The analytic results are given by eq. (10.6). The values have been normalized by the maximum value of the Alba-Calabrese formula. The UV-cutoff is set to $\epsilon = 1$.

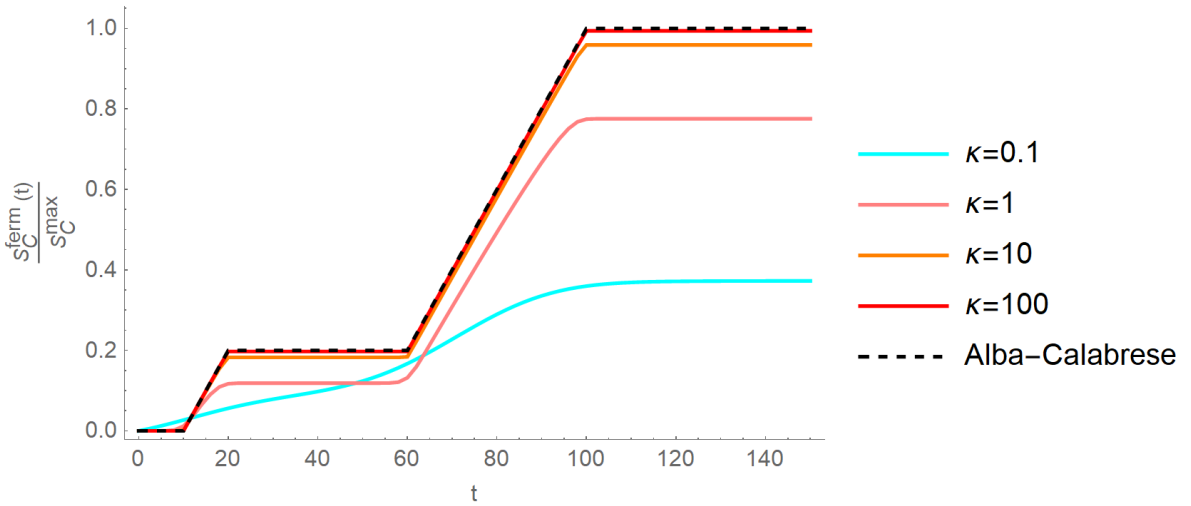


Figure 11: The entanglement entropy for a free massless fermion field on the interval $C = [10, 20] \cup [60, 100]$ for different values of the temperature parameter $\kappa$, compared to the corresponding Alba-Calabrese formula. The values have been normalized by the maximum value of the Alba-Calabrese formula. The UV-cutoff is set to $\epsilon = 1$.

The result is exactly analogous to the linear part of eq. (10.4), with $\lambda$ playing the role of temperature and $N$ playing the role of the central charge. The variables $\kappa$ and $\lambda$ also give equivalent thermal fluxes of $\frac{c\kappa^2}{48\pi}$ and $\frac{N\lambda^2}{48\pi}$ (in the latter case we use the standard normalization of the stress tensor, and not the $N$ dependent normalization of section 7.1). It would be interesting to investigate further the validity of the Alba-Calabrese formula in the RST-model.

# Part IV

# Conclusions

## 11 Summary and Outlook

In this thesis, we first reviewed recent progress on the black hole information paradox. The first key element of the story was the formulation of the generalized entropy based on the AdS/CFT correspondence [39] (section 6.2). The generalized entropy gives the entropy of a spacetime region in terms of a semi-classical quantum contribution and the area of a surface in spacetime. The surface is such that it extremizes the sum of its area and the quantum contribution, and is known as a quantum extremal surface. The second major step was then to apply the generalized entropy to more general cases of quantum fields coupled to gravity. Importantly, for an evaporating black hole, a quantum extremal surface appears near the event horizon. The resulting generalized entropy reproduces the behaviour expected of a unitary theory of quantum gravity, illustrated by the Page curve (see section 5). We show this explicitly for the RST-model in section 7.1, which provides an analytically tractable example of a two-dimensional black hole.

We also investigated the generalized entropy of a black hole before the appearance of the quantum extremal island, and used the RST model to show that the generalized entropy of the interior can grow larger than the bound set by the Bekenstein-Hawking entropy before the appearance of a non-empty quantum extremal surface (section 8). Thus, we are faced with the puzzling situation where the semi-classical description can give results that conflict with unitarity when applied to the black hole interior, and yet the approximation must be applied in the interior to compute the generalized entropy. A potential way to avoid this issue could be to avoid the use of time-slices that penetrate far into the interior of the black hole. Before the quantum extremal surface appears near the horizon, such a slice would have to extend back in time, to the past of the formation of the black hole. If this is possible, it would allow one to obtain the Page curve for a region containing a black hole, while remaining agnostic about the degrees of freedom in the interior. This would be an interesting direction for further investigation.

In addition, we applied an approximation scheme known from condensed matter physics

to a moving mirror model (see section 10). We tested the result against the example case of a free massless fermion, and found that the approximation gives accurate results for thermal radiation at high temperature. For an RST black hole of large mass, an observer far from the black hole (near future null infinity) sees a thermal spectrum (see section 7.1), and accordingly the entropy behaves similarly to what is expected from the quasiparticle picture (see equation 10.9). It would be interesting to search for a more detailed quasiparticle description of black hole evaporation in the RST model.

By reproducing the Page curve, the generalized entropy states that the Hawking radiation emitted at late times is correlated with the early radiation. This would resolve not only Page's paradox, but also Hawking's formulation of the paradox (as the late stage evaporation would approach a pure state). However, the generalized entropy does not provide a mechanism by which these correlations arise, and to the best of our knowledge no such mechanism is apparent in the the models that have been investigated thus far. Identifying such a mechanism is a clear long term goal to guide further work on this topic.

Derivations of the Page curve usually provide an exact QFT computation for the entanglement entropy of matter fields, which sometimes comes at the cost of having to use black hole models that differ from proper four-dimensional black holes in ways that could be significant. For example, in the RST-model the black hole does not have curvature singularity. The same is true in Jackiw-Teitelboim gravity (see for example [52]), which is another widely studied dilaton theory in the context of quantum extremal islands. If possible, it would therefore be interesting to study the generalized entropy of a four-dimensional evaporating black hole with a curvature singularity, perhaps using a very simple qubit model for the radiation along the lines of [12].

# A Conformal Field Theory

In order to make explicit calculations in this thesis, it will be useful to restrict ourselves to quantum field theories with strong symmetry-constraints which fix several properties of the theory. In particular, we will concern ourselves with conformal field theories (CFT), which is the class of theories for which the action is invariant under all transformations $x^\alpha \to \tilde{x}^\alpha$ that satisfy:

$$g_{\alpha\beta} \to \tilde{g}_{\alpha\beta} = e^{-2\rho} g_{\alpha\beta}, \tag{A.10}$$

where $\rho$ is a real scalar function. As is clear from this definition, the presence of conformal symmetry implies that the classical theory is not sensitive to (local) changes of the length-scale, as long as angles are kept fixed. To avoid giving an overly detailed introduction to CFT's, we will merely state the general properties that are most relevant for our purpose. For a more detailed introduction to this vast topic we refer the reader to the references [53, 54], on which much of our brief introduction is based.

## A.1 Toolbox

The restrictive assumption of conformal symmetry allows one to make powerful statements about the entire class of CFTs. For our purpose, these statements will typically be transformation laws showing how certain objects behave under conformal transformations. It should be noted that the class of CFTs is still general enough to contain theories with a large variety of behaviours, including both free and interacting theories.

In this thesis we will mainly concern ourselves with calculations in two dimensions, where the discussion of conformal transformations is particularly elegant. Specifically, if we introduce the complex coordinates $(u = x + it, v = -x + it)$ in the Euclidean plane[31,32], then the transformation $(u, v) \to (U(u), V(v))$ is conformal if and only if the functions $U$ and $V$ are holomorphic. One way to see this is to note that a general holomorphic transformations is locally approximated by a translation, a dilation and a rotation, all of which are angle-preserving.

**Primary operators:**

On a classical level, CFTs transform in an intuitive way under conformal transformations. For example, the $u$-component of a vector will transform as $\psi_u(u) \to U'(u)\psi_u(U(u))$, which is just the ordinary vector transformation law written in complex notation. More generally, for a component of a rank $N$ tensor we get

$$\psi_{u^a v^b}(u, v) \to (U'(u))^a (V'(v))^b \, \psi_{u^a v^b}(U(u), V(v)), \tag{A.11}$$

where $a$ and $b$ are the number of $u$ and $v$ indexes respectively, with $a + b = N$. An operator is called *primary* if it shows a similar transformation behaviour after the theory has been quantized. More specifically, primary operators obey

$$\psi(u, v) \to (U'(u))^h (V'(v))^{\tilde{h}} \, \psi(U(u), V(v)), \tag{A.12}$$

where the theory specific real numbers $(h, \tilde{h})$ are known as conformal weights. Note that the conformal weights are not necessarily equal to the classical exponents $(a, b)$, and are in general not integers.

One reason we are interested in primary operators is that one can deduce some general properties from just knowing the conformal weights. As an example, consider a two point function of primary operators. Applying eq. (A.12) for the specific example of a dilation $(u, v) \to (\lambda u, \lambda v)$, we get

$$\langle \psi(u, v)\psi(u', v') \rangle = \lambda^{2(h+\tilde{h})} \langle \psi(\lambda u, \lambda v)\psi(\lambda u', \lambda v') \rangle. \tag{A.13}$$

[31] Euclidean here meaning that the metric is flat and positive definite, as in conventional Euclidean geometry. Although we use Euclidean signature here for simplicity, the results below are still applicable in Lorentzian signature.

[32] To be clear, although it looks like $u$ is just the conjugate of $-v$, one can think of the two as independent complex coordinates when doing calculations. One then sets $u$ equal to the conjugate of $-v$ at the end of the calculation, which gives the physical result.

The two point function must also have translational and rotational invariance, meaning that it can only depend on $|u-u'||v-v'|$. Combining these two facts fixes the form of the two-point function, giving

$$\langle\psi(u,v)\psi(u',v')\rangle = \frac{C}{(|u-u'||v-v'|)^{h+\bar{h}}}, \tag{A.14}$$

where $C$ is a constant that does not concern us here. In section 6.1, we use this result to establish a general formula for the entropy of conformal fields.

It will also be be useful to have an expressions for Weyl-transformations, which are conformal rescalings of the metric on the form $ds^2 = dudv \to ds^2 = e^{-2\rho(u,v)}dudv$. In this case, assuming $h = \tilde{h}$, we have

$$\langle\psi(u,v)\rangle \to e^{2h\rho(u,v)}\langle\psi(u,v)\rangle. \tag{A.15}$$

This can be used to generalize the vacuum entropy formula eq. (6.10) to the curved space, as in (7.18).

**Stress tensor transformation law:**

The stress tensor is an (non-primary) operator of central importance in conformal field theory. In this thesis, we need it because it acts as a source for the equations of motion in section 7.1. The most important property for our purpose is its behaviour under conformal coordinate transformations, which takes the form

$$\langle T_{uu}(u,v)\rangle \to U'(u)^{-2}\left(\langle T_{uu}(u,v)\rangle - \frac{c}{24\pi}\{U(u),u\}\right), \tag{A.16}$$

where

$$\{U(u),u\} = \frac{U'''(u)}{U'(u)} - \frac{3}{2}\left(\frac{U''(u)}{U'(u)}\right)^2 \tag{A.17}$$

is known as the Schwarzian derivative, and theory-specific constant $c$ is called the central charge. An equivalent expression holds for the $vv$-component. The second term of the transformation law is a normal ordering correction due to the different definitions of the vacuum in the two sets of coordinates.

# B Penrose Diagrams

A Penrose diagram is a useful tool for visualizing the causal structure of a spacetime. Essentially, the diagram is a compact two-dimensional drawing of a spherically symmetric spacetime in which angles are faithfully represented. As we shall see, they are particularly useful for understanding properties of black holes. For an introduction to Penrose diagrams, see e.g. [55].

*Minkowski space:*

To start, consider flat Minkowski space in spherical coordinates:

$$ds^2 = -dt^2 + dr^2 + r^2 d\Omega^2, \tag{B.18}$$

where $\Omega$ is the spherical line element. Because of the spherical symmetry, we can restrict our focus to the $r$ and $t$ directions. Changing to a set of compact light-cone coordinates

$$u = \arctan(t - r), \tag{B.19a}$$

$$v = \arctan(t + r), \tag{B.19b}$$

we get

$$ds^2 = -e^{-2\rho} du dv, \tag{B.20}$$

where $\rho = \ln(\cos(u)\cos(v))$. The new metric is just the flat Minkowski metric rescaled by a factor of $e^{-2\rho}$. Such rescalings are *conformal*, meaning that they leave angles invariant.

We can now draw a representation of Minkowski space in the $u, v$ coordinates, see Fig. 12. Due to the singularities of the conformal factor, points on the boundary will represent asymptotic regions of the spacetime. In particular, two point labeled $i^0$ represent the region that is infinitely spacelike separated from the origin, $i^{\pm}$ represents regions that are infinitely timelike separated from the origin, and the edges $\mathcal{I}^{\pm}$ represent regions that are lightlike separated from the origin.

Importantly, diagonal lines in Fig. 12 represent lightlike trajectories in the $t$-$r$ plane, just like in the standard Minkowski coordinates. Indeed, this is a general feature of Penrose diagrams, which makes them a useful tool for visualizing the causal structure of a given spacetime.

*Schwarzschild black hole:*

Here we briefly outline the standard derivation of the Penrose diagram for the Schwarzschild black hole in case the reader wishes to better understand the diagrams in Fig. 1. More details can be found in [55]. First, we put the $t - r$ components of the metric on a conformally flat form by changing the radial coordinate to

$$r_* = r + r_s \ln\left(\frac{r}{r_s} - 1\right), \tag{B.21}$$

where $r > r_s \equiv 2G_N M$. In terms of the light cone coordinates

$$u' = t - r_*, \qquad v' = t + r_*, \tag{B.22}$$

we get

$$ds^2 = -\left(1 - \frac{r_s}{r}\right) du' dv', \tag{B.23}$$

where $r$ is now a function of $u'$ and $v'$. The coordinate singularity at $r = r_s$ can be removed by transforming to Kruskal coordinates

$$U = -\exp\left(-\frac{u'}{2r_s}\right), \qquad V = \exp\left(\frac{v'}{2r_s}\right), \tag{B.24}$$

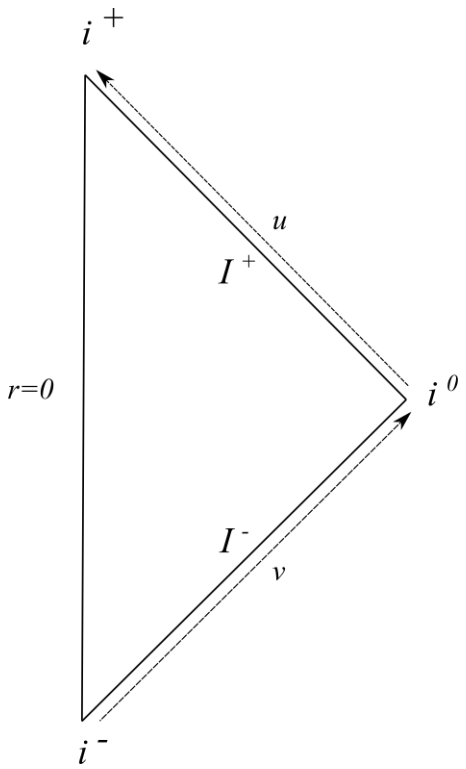


Figure 12: A Penrose diagram of flat Minkowski space showing the radial and time directions. The diagram can also be viewed as the right half of a two dimensional Minkowski plane, in which case allowing $r < 0$ would give a diagram for the full spacetime.

giving the metric

$$ds^2 = -\left(\frac{4r_s^3}{r}e^{-r/r_s}\right) dU dV, \tag{B.25}$$

which only diverges at the physical singularity at $r = 0$. We imposed $r > r_s$, so the coordinates in eq. (B.24) only cover the exterior of the black hole. However, we could have restricted ourselves to the interior instead by flipping the sign inside the logarithm in eq. (B.21). This would give a corresponding change of sign for $U$ in eq. (B.24). Thus, we can extend the Kruskal coordinates to the interior by extending the range of $U$ to $-\infty < U < \infty$. The Maximally extended solution in Fig. 1 is then obtained by also extending the range of $V$ to $-\infty < V < \infty$, which gives rise to two new regions not described by the Schwarzschild coordinates. One corresponds to the second exterior region in the left diagram in Fig. 1, and the other is the past interior region containing the past singularity. Finally, we transform to the compact light cone coordinates

$$\tilde{U} = \arctan(U), \qquad \tilde{V} = \arctan(V), \tag{B.26}$$

which gives us the left Penrose diagram in Fig. 1. For the right Penrose diagram in Fig. 1 where a Schwarzschild black hole forms from the collapse of a null shock wave, we have

flat Minkowski space to the past of the shock wave and the Schwarzschild geometry to the future of the shock wave, so the Penrose diagram can be obtained by appropriately patching together the vacuum diagram Fig. 12 and the maximally extended Schwarzschild Penrose diagram, see [55].

# References


[1] S. W. Hawking, "Particle Creation by Black Holes," *Commun. Math. Phys.* **43** (1975) 199–220. [Erratum: Commun.Math.Phys. 46, 206 (1976)].

[2] S. W. Hawking, "Breakdown of Predictability in Gravitational Collapse," *Phys. Rev. D* **14** (1976) 2460–2473.

[3] D. N. Page, "Information in black hole radiation," *Phys. Rev. Lett.* **71** (1993) 3743–3746, arXiv:hep-th/9306083.

[4] D. N. Page, "Time Dependence of Hawking Radiation Entropy," *JCAP* **09** (2013) 028, arXiv:1301.4995 [hep-th].

[5] A. Almheiri, N. Engelhardt, D. Marolf, and H. Maxfield, "The entropy of bulk quantum fields and the entanglement wedge of an evaporating black hole," *JHEP* **12** (2019) 063, arXiv:1905.08762 [hep-th].

[6] G. Penington, "Entanglement Wedge Reconstruction and the Information Paradox," *JHEP* **09** (2020) 002, arXiv:1905.08255 [hep-th].

[7] C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation*. W. H. Freeman, San Francisco, (1973).

[8] T. Jacobson, "Introductory lectures on black hole thermodynamics," (1996).

[9] J. D. Bekenstein, "Black holes and the second law," *Lett. Nuovo Cim.* **4** (1972) 737–740.

[10] A. Strominger and C. Vafa, "Microscopic origin of the Bekenstein-Hawking entropy," *Phys. Lett. B* **379** (1996) 99–104, arXiv:hep-th/9601029.

[11] T. Jacobson, "Renormalization and black hole entropy in Loop Quantum Gravity," *Class. Quant. Grav.* **24** (2007) 4875–4879, arXiv:0707.4026 [gr-qc].

[12] S. D. Mathur, "The Information paradox: A Pedagogical introduction," *Class. Quant. Grav.* **26** (2009) 224001, arXiv:0909.1038 [hep-th].

[13] T. Jacobson, "Introduction to quantum fields in curved space-time and the Hawking effect," in *School on Quantum Gravity*, pp. 39–89. 8, (2003). arXiv:gr-qc/0308048.

[14] M. E. Peskin and D. V. Schroeder, *An Introduction to quantum eld theory*. Addison-Wesley, Reading, USA, (1995).

[15] E. Frodden and N. Valdés, "Unruh Effect: Introductory Notes to Quantum Effects for Accelerated Observers," *Int. J. Mod. Phys. A* **33** no. 27, (2018) 1830026, arXiv:1806.11157 [gr-qc].

[16] N. D. Birrell and P. C. W. Davies, *Quantum Fields in Curved Space*. Cambridge Monographs on Mathematical Physics. Cambridge University Press, (1982).

[17] M. R. R. Good, E. V. Linder, and F. Wilczek, "Moving mirror model for quasithermal radiation fields," *Phys. Rev. D* **101** no. 2, (2020) 025012, arXiv:1909.01129 [gr-qc].

[18] P. C. W. Davies and S. A. Fulling, "Radiation from a moving mirror in two-dimensional space-time conformal anomaly," *Proc. Roy. Soc. Lond. A* **348** (1976) 393–414.

[19] W. R. Walker, "Particle and Energy Creation by Moving Mirrors," *Phys. Rev. D* **31** (1985) 767–774.

[20] T. P. Singh and T. Padmanabhan, "Notes on semiclassical gravity," *Annals Phys.* **196** (1989) 296–344.

[21] H. Casini, M. Huerta, and J. A. Rosabal, "Remarks on entanglement entropy for gauge fields," *Phys. Rev. D* **89** no. 8, (2014) 085012, arXiv:1312.1183 [hep-th].

[22] A. Almheiri, T. Hartman, J. Maldacena, E. Shaghoulian, and A. Tajdini, "The entropy of Hawking radiation," (2020) , arXiv:2006.06872 [hep-th].

[23] V. E. Hubeny, M. Rangamani, and T. Takayanagi, "A Covariant holographic entanglement entropy proposal," *JHEP* **07** (2007) 062, arXiv:0705.0016 [hep-th].

[24] K. Skenderis and B. C. van Rees, "Real-time gauge/gravity duality: Prescription, Renormalization and Examples," *JHEP* **05** (2009) 085, arXiv:0812.2909 [hep-th].

[25] M. Rangamani and T. Takayanagi, *Holographic Entanglement Entropy*, vol. 931. Springer, (2017). arXiv:1609.01287 [hep-th].

[26] M. R. Zirnbauer, "Another critique of the replica trick," (1999) , arXiv:cond-mat/9903338.

[27] H. Casini and M. Huerta, "Entanglement entropy in free quantum field theory," *J. Phys. A* **42** (2009) 504007, arXiv:0905.2562 [hep-th].

[28] P. Calabrese and J. L. Cardy, "Entanglement entropy and quantum field theory," *J. Stat. Mech.* **0406** (2004) P06002, arXiv:hep-th/0405152.

[29] J. M. Maldacena, "The Large N limit of superconformal field theories and supergravity," *Adv. Theor. Math. Phys.* **2** (1998) 231–252, arXiv:hep-th/9711200.

[30] E. Witten, "Anti-de Sitter space and holography," *Adv. Theor. Math. Phys.* **2** (1998) 253–291, arXiv:hep-th/9802150.

[31] S. S. Gubser, I. R. Klebanov, and A. M. Polyakov, "Gauge theory correlators from noncritical string theory," *Phys. Lett. B* **428** (1998) 105–114, arXiv:hep-th/9802109.

[32] S. El-Showk and K. Papadodimas, "Emergent Spacetime and Holographic CFTs," *JHEP* **10** (2012) 106, arXiv:1101.4163 [hep-th].

[33] S. Ryu and T. Takayanagi, "Aspects of Holographic Entanglement Entropy," *JHEP* **08** (2006) 045, arXiv:hep-th/0605073.

[34] A. Lewkowycz and J. Maldacena, "Generalized gravitational entropy," *JHEP* **08** (2013) 090, arXiv:1304.4926 [hep-th].

[35] H. Goldstein, C. Poole, and J. Safko, *Classical Mechanics (3rd Edition)*. (2001).

[36] X. Dong, A. Lewkowycz, and M. Rangamani, "Deriving covariant holographic entanglement," *JHEP* **11** (2016) 028, arXiv:1607.07506 [hep-th].

[37] S. N. Solodukhin, "Entanglement entropy of black holes," *Living Rev. Rel.* **14** (2011) 8, arXiv:1104.3712 [hep-th].

[38] T. Faulkner, A. Lewkowycz, and J. Maldacena, "Quantum corrections to holographic entanglement entropy," *JHEP* **11** (2013) 074, arXiv:1307.2892 [hep-th].

[39] N. Engelhardt and A. C. Wall, "Quantum Extremal Surfaces: Holographic Entanglement Entropy beyond the Classical Regime," *JHEP* **01** (2015) 073, arXiv:1408.3203 [hep-th].

[40] T. Hartman, E. Shaghoulian, and A. Strominger, "Islands in Asymptotically Flat 2D Gravity," *JHEP* **07** (2020) 022, arXiv:2004.13857 [hep-th].

[41] J. G. Russo, L. Susskind, and L. Thorlacius, "The Endpoint of Hawking radiation," *Phys. Rev. D* **46** (1992) 3444–3449, arXiv:hep-th/9206070.

[42] T. M. Fiola, J. Preskill, A. Strominger, and S. P. Trivedi, "Black hole thermodynamics and information loss in two-dimensions," *Phys. Rev. D* **50** (1994) 3987–4014, arXiv:hep-th/9403137.

[43] D. Grumiller, W. Kummer, and D. V. Vassilevich, "Dilaton gravity in two-dimensions," *Phys. Rept.* **369** (2002) 327–430, arXiv:hep-th/0204253.

[44] P. Calabrese, J. Cardy, and E. Tonni, "Entanglement entropy of two disjoint intervals in conformal field theory," *J. Stat. Mech.* **0911** (2009) P11001, arXiv:0905.2069 [hep-th].

[45] M. Varadarajan, "A Note on Entanglement Entropy, Coherent States and Gravity," *Gen. Rel. Grav.* **48** no. 3, (2016) 35, arXiv:1602.00106 [gr-qc].

[46] I. Booth, "Black hole boundaries," *Can. J. Phys.* **83** (2005) 1073–1099, arXiv:gr-qc/0508107.

[47] V. Alba and P. Calabrese, "Entanglement and thermodynamics after a quantum quench in integrable systems," *Proceedings of the National Academy of Science* **114** no. 30, (July, 2017) 7947–7951, arXiv:1608.00614 [cond-mat.str-el].

[48] V. Alba and P. Calabrese, "Entanglement dynamics after quantum quenches in generic integrable systems," *SciPost Phys.* **4** no. 3, (2018) 017, arXiv:1712.07529 [cond-mat.stat-mech].

[49] S. Chapman, J. Eisert, L. Hackl, M. P. Heller, R. Jefferson, H. Marrochio, and R. C. Myers, "Complexity and entanglement for thermofield double states," *SciPost Phys.* **6** no. 3, (2019) 034, arXiv:1810.05151 [hep-th].

[50] I. Akal, Y. Kusuki, N. Shiba, T. Takayanagi, and Z. Wei, "Entanglement Entropy in a Holographic Moving Mirror and the Page Curve," *Phys. Rev. Lett.* **126** no. 6, (2021) 061604, arXiv:2011.12005 [hep-th].

[51] I. A. Reyes, "Moving mirrors, Page curves and bulk entropies in $AdS_2$," (2021) , arXiv:2103.01230 [hep-th].

[52] A. Almheiri and J. Polchinski, "Models of $AdS_2$ backreaction and holography," *JHEP* **11** (2015) 014, arXiv:1402.6334 [hep-th].

[53] P. D. Francesco, P. Mathieu, and D. Sénéchal, *Conformal Field Theory*. Springer New York, (1997).

[54] A. N. Schellekens, "Introduction to conformal field theory," *Fortsch. Phys.* **44** (1996) 605–705.

[55] A. Strominger, "Les Houches lectures on black holes," in *NATO Advanced Study Institute: Les Houches Summer School, Session 62: Fluctuating Geometries in Statistical Mechanics and Field Theory*. (1994). arXiv:hep-th/9501071.