Swansea University

Swansea University
Prifysgol Abertawe

The black hole information problem and JT gravity

Neil Talwar

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

in the

Department of Physics,
Faculty of Science and Engineering,
Swansea University.

August 2024

Declaration of Authorship

I, Neil Talwar, declare that this thesis titled, "The black hole information problem and JT gravity" and the work presented in it are my own. I confirm that:

- This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.
- This thesis is the result of my own investigations, except where otherwise stated. Other sources are acknowledged by footnotes giving explicit references. A bibliography is appended.
- I hereby give consent for my thesis, if accepted, to be available for photocopying and for interlibrary loan, and for the title and summary to be made available to outside organisations.
- The University's ethical procedures have been followed and, where appropriate, that ethical approval has been granted.

Signed: N. Talwar
Date: 4/08/24

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my supervisor, Tim Hollowood. His continuous support and guidance throughout these past four years have been truly invaluable. I thank him for always making the time to patiently answer all my questions and for encouraging me to become more independent.

I would also like to acknowledge my collaborators. I am deeply grateful for the opportunity to work with Prem Kumar; his generosity with time has been immensely helpful, and his passion for his work has been truly inspiring. It has also been a pleasure working with Andrea Legramandi, who has always shown great patience with my continuous questioning of our results. I have learned a great deal through our many discussions, whether in person, over Zoom, or through countless Discord messages. Additionally, it has been fun to work with Zsolt Gyongyosi. I have always enjoyed our discussions and his attempts to explain the formal aspects of operator algebras to me.

This work would have not been possible without the companionship of my fellow students in room 666 of Vivian Tower. I have learnt a lot of physics through our discussions, and all the dinners, drinks, climbing, and hiking has made the experience so much fun. So, I'd like to thank: Lewis, Ricardo, Emily, David, Mohammed, Lucas, Chanju, Ali, Dimitris, Federico, Zsolt, Luke, Karol, Natalia, Ameek, Shaun, and Nick.

I also cannot forget to thank Adam, Tanuj, James, and the 6 guys for being such greats friends.

Most importantly, I would like to thank my mum and dad, and my sister Sara, for their unwavering support. Without them, none of this would have been possible.

Contents

Chapter 1

Introduction

General relativity tells us that, quite generically, compact objects, such as heavy stars, can and will undergo gravitational collapse to form black holes [1]. Black holes are characterised by an event horizon, a surface which divides spacetime into two parts: the exterior of the black hole and the interior of the black hole, a region from which nothing, not even light, can escape. As a consequence, for an observer in the exterior region, there is apparently no way to get a signal from and learn anything about the interior.

This peculiar feature of black holes led Bekenstein [2] to assign a black hole an entropy proportional to its horizon area in an attempt to save the second law of thermodynamics, which could otherwise be violated by throwing matter into a black hole. This proposal has an obvious flaw: if black holes have entropy then they must generically have a nonzero temperature. Yet, by definition, classical black holes have zero temperature since they cannot emit anything. This dilemma was later resolved by Hawking [3], who considered the behaviour of quantum fields on a black hole spacetime and showed that black holes do in fact emit radiation with a perfect thermal spectrum. By the first law of thermodynamics, we can then associate a classical thermodynamic entropy to the black hole, known as the Bekenstein-Hawking entropy,

$$S_{\mathrm{BH}} = \frac{A}{4G}, \tag{1.0.1}$$

where A is the horizon area. This entropy is remarkably large. For instance, a solar mass black hole has $S_{\mathrm{BH}} \sim 10^{77}$. Although black holes, in some respects, seem to behave like ordinary objects that obey the laws of the thermodynamics, Hawking [4] discovered that they seem to differ in a fundamental way. The radiation emitted by a black hole is noisy; it has a perfect thermal spectrum and is seemingly unrelated to the state of the matter that formed the black hole. This is in tension with unitarity evolution in quantum mechanics if the radiation remains in a perfect thermal state, even after the black hole has entirely evaporated. Information would then be truly lost.

A sharp formulation of this puzzle is due to Page [5], who used the entropy of the radiation as a diagnostic for information loss. Hawking's calculation suggests that the entropy of the radiation keeps rising as the black hole evaporates and emits radiation[1]. This follows because the emission of Hawking radiation is dominated by the s-wave sector of massless fields which can be approximated by a 1+1-dimensional gas of bosons or fermions. This reasoning leads to an entropy which increases linearly in time. The intuitive reason behind this growing entropy is the stretching of space behind the horizon which can be understood using the "nice slice" picture of [6, 7]. Eventually, we run into a contradiction at the Page time, the point at which the entropy becomes larger than the Bekenstein-Hawking entropy. If we believe that the Bekenstein-Hawking entropy has a statistical mechanical underpinning, then at this point, there are simply not enough states to purify all the radiation[2]. At this point, the entropy must begin to decrease, following what is known as the Page curve. This sharpens the argument because the puzzle now shows up a lot sooner, when the black hole is roughly half its initial size. This argument reveals a fundamental tension between the evaporation of black holes in semiclassical gravity and the unitarity of quantum mechanics.

Recent developments [9, 10] in the black hole information problem have pointed to a solution requiring only the semiclassical toolkit. The approach relies on the inclusion of spacetime wormholes in the gravitational path integral when computing the entropy using the replica trick [11, 12]. After the Page time, the dominant saddle point for computing the entropy involves a wormhole connecting the replicas. In fact, the calculation is conceptually similar to the derivation of the Bekenstein-Hawking entropy from the gravitational path integral by Gibbons and Hawking [13]. The basic role of wormhole contributions is to provide nonperturbatively small corrections to the naive bulk inner product for the modes inside the horizon with which the radiation is entangled. These small corrections gradually build up over time and eventually have a dramatic effect at the Page time. This resonates with the widely held belief that information is not really lost but instead ends up being encoded in subtle correlations in the radiation.

These developments grew out of ideas in the context of the AdS/CFT correspondence [14–16], starting with the work of Ryu and Takayanagi [17], who showed that the Bekenstein-Hawking entropy formula (1.0.1) is a special case of a more general formula for computing the entropy of boundary subregions B in the CFT,

$$S(B) = \frac{A(\gamma_B)}{4G}, \tag{1.0.2}$$

where $A(\gamma_B)$ is the area of a minimal surface γ_B in the bulk which ends on the boundary of the subregion. The formula was soon revised to include time dependence [18] and

[1] Mathur's "small corrections theorem" [6] demonstrates that small corrections to the state of the Hawking radiation cannot change this result. However, recent work has shown how this result can change once certain nonperturbative effects are taken into account.

[2] This idea—that a black hole, when viewed from the outside, can be described in terms of a quantum system with $e^{S_{\text{BH}}}$ states—is sometimes called the *central dogma* [8].

quantum corrections [19, 20],

$$S(B) = \frac{A(\gamma_B)}{4G} + S_{\text{bulk}}(b), \tag{1.0.3}$$

where the surface γ_B, known as the quantum extremal surface (QES), is determined by extremising the quantity on the RHS and $S_{\text{bulk}}(b)$ is the entropy of bulk quantum fields on a partial Cauchy surface b in the bulk such that $\partial b = \gamma_B \cup B$. In fact, the QES prescription turns out to be a signature of a deeper relation between the bulk and boundary known as entanglement wedge reconstruction[3] [19, 24–29], which states that the boundary subregion B has access to all the information in the bulk contained within its entanglement wedge b. The idea of entanglement wedge reconstruction is perhaps most naturally understood in the language of quantum error correction, where the linear map from the semiclassical bulk Hilbert space to the Hilbert space of the CFT is the error correcting code [30, 31].

Extending these ideas to evaporating black holes provides a window into how information eventually gets encoded in the radiation. The initial breakthrough was driven by the discovery of novel QESs which can contribute to entropy of Hawking radiation and give rise to the Page curve [9, 10]. More generally, the QES prescription allows one to understand correlations between subsets of the Hawking radiation. There are also ongoing developments into understanding how measures of multipartite entanglement [32, 33] can give further insight into the entanglement structure of the radiation.

More generally, spacetime wormholes have been crucial in understanding how the discrete nature of black hole microstates can be analysed using the semiclassical toolkit. Beyond the Page curve, they also explain some aspects of the spectral form factor [34, 35], and the behaviour of late-time correlation functions [36–39]. The latter lies at the heart of a version of the information problem for an eternal black hole in AdS due to Maldacena [40]. In gravity, the two-point function of quantum fields outside the horizon of an eternal black hole appears to decay to zero as the fields are separated further and further in time. On the boundary, the two-point function is just that of a quantum mechanical system with a discrete spectrum in a thermal state,

$$\langle \mathcal{O}(t)\mathcal{O}(0)\rangle_\beta = \frac{1}{Z(\beta)} \sum_{m,n=1}^{D} |\mathcal{O}_{mn}|^2 e^{-\beta E_m - \mathrm{i}t(E_m - E_n)}. \tag{1.0.4}$$

The discreteness of the spectrum forbids the correlator from decaying in time indefinitely, instead the phases cause the correlator to rattle around a small but nonzero value. This conflict is often referred to as Maldacena's version of the information problem [40] and is particularly sharp since it can be phrased entirely within the context of the AdS/CFT correspondence.

[3] For a discussion on the subtleties and the scope of validity of entanglement wedge reconstruction, see [21–23].

The two-point function at late times is an erratic function which depends strongly on the details of the energy spectrum. However, by studying a coarse-grained version of this quantity—such as by taking a running time average—and assuming that the Hamiltonian exhibits random matrix universality and that the matrix elements of the operators obey the eigenstate thermalization hypothesis [41, 42], a universal structure emerges. After decaying for a while, the mean signal of the two-point function follows a linear ramp before saturating on a plateau.

Contributions of spacetime wormholes to the gravitational path integral are best understood in simple models, most notably Jackiw-Teitelboim (JT) gravity [43, 44]. This two-dimensional model of gravity came into the spotlight when Kitaev [45] showed that it emerges as the effective low-energy description of a disordered quantum mechanical system known as the Sachdev-Ye-Kitaev (SYK) model [45, 46]. This is a version of holographic duality with a crucial difference: JT gravity is dual to an ensemble of quantum systems, not to a specific quantum system. In this duality, the role of the gravitational path integral gains clarity: it explains statistical aspects of black hole microstates. For instance, its explains some aspects of the mean value of the correlator (1.0.4) and the statistics of the late-time noise, but not the noise itself.

The plan of this thesis is as follows. We start in chapter 2 by discussing some aspects of JT gravity coupled to a matter quantum field theory. We review how evaporating black holes can be formed in JT gravity by coupling to an auxiliary system. The aim of this chapter is to introduce some background material and notation which will be used throughout chapters 3-5.

In chapter 3 we analyse the entropy of subsets of Hawking radiation for an evaporating black hole in JT gravity. We identify a simple class of QESs and show how the entropy is determined by a nontrivial competition between multiple QESs. We also extend this analysis to the case with an interface between the gravitational region and the auxiliary system, which serves as a toy model for a black hole with greybody factors. We compare these findings with a microscopic toy model that accounts for thermodynamic irreversibility of evaporation due to greybody factors.

In chapter 4 we examine when information which has been thrown into an evaporating black hole can be recovered in the radiation, a problem first studied by Hayden and Preskill [47]. We solve exactly for backreaction of the infalling object on the geometry within the JT model and examine how the dimension of the infalling system affects the time scale for the information to be recoverable in the radiation. We compare these results with a microscopic toy model in which we study reconstruction of the ingoing modes using the Petz map.

From a holographic perspective, JT gravity coupled to an auxiliary system has a dual description as a quantum mechanical system coupled the auxiliary system. In an appropriate infrared limit, boundary conformal field theory (BCFT) offers a powerful framework to examine this setup, which we explore in section 5. At high temperatures,

the dynamics of entropies in BCFT are controlled by certain OPE limits of twist operator correlation functions used to compute the Rényi entropy. We show how this behaviour is connected to the behaviour of certain OPE limits of twist operators in the bulk matter theory, which can involve contributions from non-trivial QESs.

In chapter 6 we study some aspects of the spectral form factor in the SYK model. Like the correlator (1.0.4), the spectral form factor is a noisy function at late times. We show how the statistics of this noise can be understood as arising from a pattern of replica symmetry breaking in the collective field variables of SYK.

Chapter 2

JT gravity

In this chapter we review some aspects of JT gravity and its coupling to a matter quantum field theory. We also discuss how evaporating black holes can be formed in JT gravity by coupling to an auxiliary system. The aim of this chapter is not to give a comprehensive introduction to this subject but rather to introduce the some background material and notation which will be used throughout chapters 3-5 of this thesis. See [48] for a recent review on this subject.

JT gravity is a simple model of gravity in two dimensions with action [43, 44]

$$I_{\rm JT} = \frac{1}{16\pi G}\left[\int_M \mathrm{d}^2x\sqrt{-g}\,\phi\,(R+2) + 2\int_{\partial M}\mathrm{d}t\sqrt{-\gamma}\,\phi\,(K-1)\right], \tag{2.0.1}$$

where g is the bulk metric with Ricci scalar R, γ is the induced metric on the boundary, and K is the extrinsic curvature of the boundary. Since the scalar field ϕ, the dilaton, appears linearly in the action its equation of motion sets $R+2=0$, which implies that the spacetime M is locally isometric to a portion of AdS_2. The last term in the action includes a Gibbons-Hawking-York boundary term and a counterterm (the -1) which is needed to cancel the divergent part of the action near an AdS_2 boundary.

Usually one also adds an Einstein-Hilbert action and Gibbons-Hawking-York boundary term

$$I_0 = \frac{\phi_0}{16\pi G}\left[\int_M \mathrm{d}^2x\sqrt{-g}\,R + 2\int_{\partial M}\mathrm{d}t\sqrt{-\gamma}\,K\right], \tag{2.0.2}$$

which, in Euclidean signature, is a topological invariant proportional to the Euler characteristic of the spacetime. The role of this term in this model will be to provide an extremal (or zero-temperature) entropy for the black hole. Remarkably, this simple model arises as the universal description of the near-horizon physics of near-extremal black holes in higher dimensions which have an approximately AdS_2 throat with a slowly varying transverse space [49]. The simplest realisation is the Reissner–Nordström black hole in four dimensions which, in the near-extremal limit, has an

approximately $\mathrm{AdS}_2 \times S^2$ near-horizon geometry. From this perspective, the area A of the transverse space is given by the dilaton

$$A = \phi_0 + \phi, \tag{2.0.3}$$

where ϕ_0 is the constant part of the dilaton which comes from the Einstein-Hilbert term (2.0.2). From this perspective, one requires $|\phi| \ll \phi_0$ to remain in the near-extremal limit.

2.1 Classical solutions

For a matter quantum field theory which does not couple directly to the dilaton ϕ, the model retains its simplicity, and the variation with respect to the dilaton and metric give the equations of motions

$$R + 2 = 0, \quad \left(-\nabla_a \nabla_b + g_{ab}\nabla^2 - g_{ab}\right)\phi = 8\pi G \langle T_{ab} \rangle. \tag{2.1.1}$$

We work in the limit where the gravitational sector may be treated semiclassically and matter quantum mechanically, so have replaced the stress tensor T_{ab} with its expectation value $\langle T_{ab} \rangle$. For simplicity we will restrict to the case that the matter is a conformal field theory (CFT). Classically, the trace of the stress tensor vanishes for a CFT. Quantum mechanically, this is only true in a flat spacetime; due to the conformal anomaly, the trace of the stress tensor is constrained to be

$$\mathrm{Tr}\, T = \frac{c}{24\pi} R, \tag{2.1.2}$$

where c is the central charge of the CFT. It will be useful to add a counterterm to cancel the source from the trace of the stress tensor. This can be achieved by adding a cosmological constant term which amounts to a redefinition of the extremal value of the dilaton $\phi_0 \to \phi_0 + \frac{cG}{3}$ [50]. Henceforth, we will assume $\mathrm{Tr}\, T = 0$. Taking the trace of the metric equation of motion (2.1.1) we then find $\nabla^2\phi = 2\phi$. This can be used to simplify the equations of motion

$$R + 2 = 0, \quad \left(-\nabla_a \nabla_b + g_{ab}\right)\phi = 8\pi G \langle T_{ab} \rangle. \tag{2.1.3}$$

2.1.1 Metric solution

Writing the metric in conformal gauge

$$\mathrm{d}s^2 = -e^{2\omega(u,v)} \mathrm{d}u \mathrm{d}v, \tag{2.1.4}$$

where u and v are lightcone coordinates, the equation of motion for the metric becomes the hyperbolic Liouville equation

$$4\partial_u \partial_v \omega + e^{2\omega} = 0. \tag{2.1.5}$$

The general solution is

$$e^{2\omega} = \frac{4\mathsf{U}'(u)\mathsf{V}'(v)}{(\mathsf{U}(u)-\mathsf{V}(v))^2}, \tag{2.1.6}$$

for some functions $\mathsf{U}(u)$ and $\mathsf{V}(v)$ of the lightcone coordinates u and v. This corresponds to the Poincaré patch of AdS_2 where the metric is

$$\mathrm{d}s^2 = -\frac{4\mathsf{U}'(u)\mathsf{V}'(v)\mathrm{d}u\mathrm{d}v}{(\mathsf{U}(u)-\mathsf{V}(v))^2} = -\frac{4\mathrm{d}\mathsf{U}\mathrm{d}\mathsf{V}}{(\mathsf{U}-\mathsf{V})^2}. \tag{2.1.7}$$

The geometry has a conformal boundary at $\mathsf{U}=\mathsf{V}$ where the metric diverges and horizons where $\mathsf{U}-\mathsf{V}=\infty$. The isometries of AdS_2 act as Möbius or fractional linear transformations on the Poincaré lightcone coordinates

$$\mathsf{U} \to \frac{\mathsf{a}\mathsf{U}+\mathsf{b}}{\mathsf{c}\mathsf{U}+\mathsf{d}}, \qquad \mathsf{V} \to \frac{\mathsf{a}\mathsf{V}+\mathsf{b}}{\mathsf{c}\mathsf{V}+\mathsf{d}}, \qquad \mathsf{a}\mathsf{d}-\mathsf{b}\mathsf{c}=1, \tag{2.1.8}$$

and form a group isomorphic to $\mathrm{PSL}(2,\mathbb{R})$. In section (2.3) we will learn how the boundary conditions of JT gravity distinguish certain coordinate frames of AdS_2, making it useful to study several important frames of AdS_2. The first frame we have encountered is the Poincaré patch (2.1.7) which only covers a subset of AdS_2. Applying the transformation

$$\mathsf{U}(t) = \mathsf{V}(t) = \tan\frac{t}{2}, \tag{2.1.9}$$

takes us to the global patch of AdS_2 where the metric is

$$\mathrm{d}s^2 = -\frac{\mathrm{d}u\mathrm{d}v}{\sin^2\frac{u-v}{2}} = \frac{-\mathrm{d}\tau^2+\mathrm{d}\sigma^2}{\sin^2\sigma}, \tag{2.1.10}$$

with $u=\tau-\sigma$ and $v=\tau+\sigma$. The global patch makes it apparent that, unlike its higher dimensional cousins, AdS_2 has a pair of conformal boundaries at $\sigma=0,\pi$. The Weyl factor $1/\sin^2\sigma$ doesn't change the causal structure of the spacetime so it's the same as that of the strip $\mathbb{R}\times(0,\pi)$ in Minkowski spacetime, which is depicted in the Penrose diagram in figure 2.1.

Another useful patch of AdS_2 is the right black hole or Rindler patch, which can be found using the transformation

$$\mathsf{U}(t) = \mathsf{V}(t) = \frac{\beta}{\pi}\tanh\frac{\pi}{\beta}t, \tag{2.1.11}$$

with parameter β. This leads to the metric

$$\mathrm{d}s^2 = -\frac{4\pi^2}{\beta^2}\frac{\mathrm{d}u\mathrm{d}v}{\sinh^2\frac{\pi}{\beta}(u-v)}. \tag{2.1.12}$$

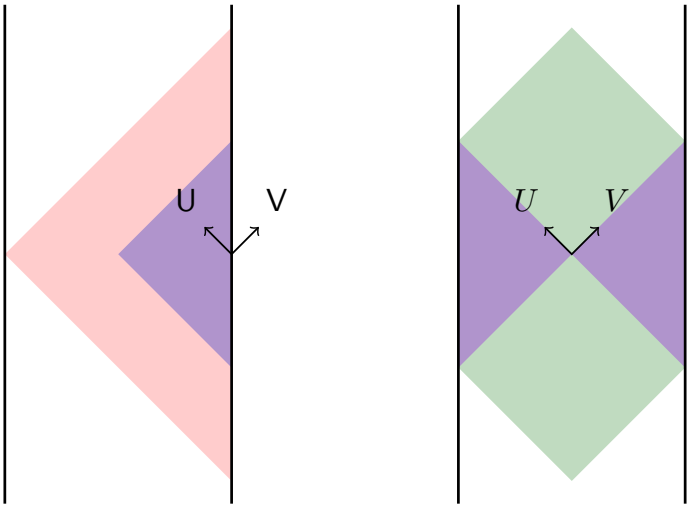


Figure 2.1: The Penrose diagram of AdS_2. The Penrose diagram on the left shows how the right black hole patch, shown in purple, is contained in the Poincaré patch, shown in pink. The Penrose diagram on the right shows how the left and right black hole patches, shown in purple, are contained in the Kruskal patch, shown in green.

The conformal boundary is at $u = v$ and there are horizons where $u - v = \infty$. This metric appears in the near-horizon limit of near-extremal black holes in higher dimensions, where the parameter β corresponds to the inverse temperature of the black hole. This becomes apparent if we introduce the coordinates

$$t = \frac{u+v}{2}, \qquad r = \frac{r_h}{\tanh \frac{\pi}{\beta}(u-v)}, \tag{2.1.13}$$

with $r_h = 2\pi/\beta$ as the metric then takes the form

$$\mathrm{d}s^2 = -(r^2 - r_h^2)\mathrm{d}t^2 + \frac{\mathrm{d}r^2}{r^2 - r_h^2}. \tag{2.1.14}$$

The identification $r_h = 2\pi/\beta$ ensures the absence of a conical singularity at the horizon when Euclidean time is made periodic with period β. For finite β the black hole patch is contained within the Poincaré patch, as shown in the left panel of figure 2.1. In the zero temperature limit $\beta \to \infty$, we recover the Poincaré patch, as $\lim_{\beta\to\infty} \frac{\beta}{\pi} \tanh \frac{\pi}{\beta} t = t$, which appears in the near-horizon limit of extremal black holes in higher dimensions.

Since the coordinates u and v do not extend to the horizons, it is often convenient to use the Kruskal coordinates

$$U = -e^{-2\pi u/\beta}, \qquad V = e^{2\pi v/\beta}. \tag{2.1.15}$$

The Kruskal coordinates extend beyond the horizon and also contain a left black hole patch, related to the right patch by the isometry $(U, V) \to (-U, -V)$. This is depicted

in the right panel of figure 2.1. The left patch can be parameterised by null coordinates $\tilde{u}$ and $\tilde{v}$ such that

$$U = e^{-2\pi\tilde{u}/\beta}, \qquad V = -e^{2\pi\tilde{v}/\beta}. \tag{2.1.16}$$

In Kruskal coordinates the metric is

$$\mathrm{d}s^2 = -\frac{4\mathrm{d}U\mathrm{d}V}{(1+UV)^2}. \tag{2.1.17}$$

Each of the Kruskal coordinates are related to the Poincaré coordinates by a Möbius transformation,

$$\mathsf{U} = -\frac{\beta}{\pi}\frac{U+1}{U-1}, \qquad \mathsf{V} = \frac{\beta}{\pi}\frac{V-1}{V+1}. \tag{2.1.18}$$

2.2 Hawking radiation from conformal anomaly

To solve the equation of motion for the dilaton (2.1.3) we need to specify the state of the matter quantum field theory. To understand which states are sensible, it is instructive to restrict the matter quantum field theory to be a CFT, where the dynamics of energy-momentum are particularly simple. For a CFT in two dimensions the trace of the stress tensor is fixed by the conformal anomaly to be proportional to the Ricci scalar (2.1.2). In lightcone coordinates, this means

$$T_{uv} = -\frac{c}{12\pi}\partial_u\partial_v\omega. \tag{2.2.1}$$

Together with conservation of energy-momentum $\nabla_a T^a{}_b = 0$, this is sufficient to determine T_{ab} up to a function of u and a function of v

$$\begin{aligned} T_{uu} &= \frac{c}{12\pi}\left(\partial_u^2\omega - (\partial_u\omega)^2\right) + t_u(u), \\ T_{vv} &= \frac{c}{12\pi}\left(\partial_v^2\omega - (\partial_v\omega)^2\right) + t_v(v). \end{aligned} \tag{2.2.2}$$

In the asymptotic region of an asymptotically AdS or flat spacetime the anomalous term vanishes and so $t_u(u)$ and $t_v(v)$ can be interpreted as the outgoing and ingoing energy flux at infinity, respectively. Specifying $t_u(u)$ and $t_v(v)$ then determines T_{ab} everywhere. Near the horizon of a static black hole $\omega \sim \frac{\pi}{\beta}(v-u)$, where β is a parameter related to the surface gravity $\kappa = 2\pi/\beta$. For the outgoing component of the stress tensor this implies

$$T_{uu} \sim -\frac{\pi c}{12\beta^2} + t_u(u). \tag{2.2.3}$$

The coordinate u does not extend to the future horizon, where $u \to \infty$, so it is convenient to use the Kruskal coordinate $U = -e^{-2\pi u/\beta}$. The Jacobian from the change of coordinates implies

$$T_{UU} \sim \frac{-\frac{\pi c}{12\beta^2} + t_u(u)}{\frac{4\pi^2}{\beta^2}U^2}. \tag{2.2.4}$$

So for any state where T_{ab} is nonsingular on the horizon $U = 0$, we must have [51]

$$t_u(u) \sim \frac{\pi c}{12\beta^2} + \mathcal{O}(e^{-4\pi u/\beta}), \quad \text{as } u \to \infty. \tag{2.2.5}$$

This is nothing but the energy density of a thermal state in a CFT if we identify β as the inverse temperature. Hence, we learn that any sensible state, with finite energy density on the horizon, necessarily has a constant thermal energy flux at inverse temperature β at late times, which is nothing but Hawking radiation. For black holes in higher dimensions the spectrum of Hawking radiation is not thermal due to the presence of greybody factors.

We also learn that any sensible state for the outgoing modes is extremely well approximated by vacuum associated to the Kruskal frame. For an evaporating black hole there is no past horizon so there is no analogous constraint for the state of the ingoing modes. For example, if there is no ingoing flux of energy from infinity $t_v = 0$, then there must be a negative flux of energy $T_{vv} = -\frac{\pi c}{12\beta^2}$ across the horizon which causes the area of the black hole to decrease. This state, where the outgoing modes are in the vacuum associated to the Kruskal frame and the ingoing modes are in the vacuum associated to the frame of an asymptotic observer, is known as the Unruh state. For example, for a massless scalar or fermion field the state for the outgoing modes in the Unruh state is

$$\bigotimes_\omega \frac{1}{\sqrt{Z_\omega}} \sum_{n=0}^{1 \text{ or } \infty} e^{-n\beta\omega/2} \, |n\rangle^\omega_{\text{int}} \otimes |n\rangle^\omega_{\text{ext}} \,, \tag{2.2.6}$$

where Z_ω is a normalisation, and $|n\rangle^\omega_{\text{int}}$ and $|n\rangle^\omega_{\text{ext}}$ denote a states of definite Rindler frequency ω in the interior and exterior, respectively. The sum goes up to 1 for a fermion and ∞ for a scalar. Tracing over interior modes gives a thermal state for the exterior outgoing modes

$$\bigotimes_\omega \frac{1}{Z_\omega} \sum_{n=0}^{1 \text{ or } \infty} e^{-n\beta\omega} |n\rangle\langle n|^\omega_{\text{ext}}. \tag{2.2.7}$$

Alternatively, the black hole can be kept in equilibrium by providing a compensating ingoing energy flux $t_v = \frac{\pi c}{12\beta^2}$. This state, where both the ingoing and outgoing modes are in the vacuum associated to the Kruskal frame, is known as the Hartle-Hawking state [52].

2.3 Boundary conditions

Let us parametrise the metric in JT gravity using Poincaré coordinates

$$\mathrm{d}s^2 = -\frac{4\mathrm{d}\mathsf{U}\mathrm{d}\mathsf{V}}{(\mathsf{U}-\mathsf{V})^2} = \frac{-\mathrm{d}F^2 + \mathrm{d}Z^2}{Z^2}, \tag{2.3.1}$$

where $\mathsf{U} = F - Z$ and $\mathsf{V} = F + Z$. In studies of JT gravity the spacetime M is usually cut off near the conformal boundary $\mathsf{U} = \mathsf{V}$ or $Z = 0$, so, M is "nearly all" of AdS_2. The physical boundary ∂M is a timelike curve $(F(t), Z(t))$ in AdS_2, which is assumed to be parameterised by a distinguished time coordinate t. To get a nearly AdS_2 spacetime one further imposes the boundary conditions

$$\mathrm{d}s^2\big|_{\partial M} = -\frac{\mathrm{d}t^2}{\epsilon^2}, \qquad \phi_b = \frac{\phi_r}{\epsilon}, \tag{2.3.2}$$

where ϕ_r is a constant and ϵ is a small parameter. For small ϵ the boundary condition for the metric implies $Z(t) = -\epsilon\, F'(t) + \mathcal{O}(\epsilon^3)$. This means that, for small ϵ, the function F entirely determines how the physical boundary ∂M is embedded in AdS_2. After imposing the negative curvature condition the action reduces to a boundary term

$$I_{\mathrm{JT}} = \frac{1}{8\pi G}\int_{\partial M} \mathrm{d}t\sqrt{-\gamma}\,\phi\,(K-1). \tag{2.3.3}$$

For small ϵ, a short calculation reveals that the trace of the extrinsic curvature is given by (see appendix A)

$$K = 1 - \epsilon^2\{F, t\} + \mathcal{O}(\epsilon^4), \tag{2.3.4}$$

where $\{F, t\}$ is the Schwarzian derivative

$$\{F, t\} = \frac{F'''(t)}{F''(t)} - \frac{3}{2}\left(\frac{F''(t)}{F'(t)}\right)^2. \tag{2.3.5}$$

Hence, the dynamics of JT gravity is encoded in a boundary action for the function F which relates the boundary time coordinate t and the Poincaré time coordinate F,

$$I_{\mathrm{JT}} = -\frac{\phi_r}{8\pi G}\int \mathrm{d}t\,\{F, t\}. \tag{2.3.6}$$

This action is known as the Schwarzian action. Actually, one gets a pair of Schwarzian actions, one for each boundary of AdS_2. The Schwarzian derivative, and hence the action, is invariant under Möbius transformations of the function F. This $\mathrm{PSL}(2,\mathbb{R})$ symmetry should thought of as a gauge symmetry which arises since not every function F gives rise to a different metric.

The equation of motion for F, found by varying the Schwarzian action, is simply energy conservation, that is, the change in energy is equal to the net influx of energy

$$\frac{\mathrm{d}E}{\mathrm{d}t} = \langle T_{vv}(t)\rangle - \langle T_{uu}(t)\rangle, \qquad E(t) = -\frac{\phi_r}{8\pi G}\{F, t\}. \tag{2.3.7}$$

The energy E is simply as defined the Noether charge associated to boundary time translations. In this expression u and v are null coordinates adapted to the boundary

region, defined by $\mathsf{U} = F(u)$ and $\mathsf{V} = F(v)$. For a given energy there is a simple method to solve

$$\{F, t\} = -\frac{8\pi G}{\phi_r} E, \tag{2.3.8}$$

for the reparameterisation function F. If f_+ and f_- are two linearly independent solutions to the linear second order ODE

$$\frac{\mathrm{d}^2 f}{\mathrm{d}t^2} = \frac{4\pi G}{\phi_r} E f, \tag{2.3.9}$$

then $F = f_+/f_-$ solves (2.3.8). The general solution is related to this one by a Möbius transformation, since two functions can only have equal Schwarzian derivative if they are related by a Möbius transformation.

2.4 Eternal black holes

A simple class of solutions correspond to equilibrium solutions, where the ingoing and outgoing energy flux are balanced. For these solutions, the equation of motion (2.3.7) imply that the energy E is a constant. The solution for F with parameter β is

$$F(t) = \frac{\beta}{\pi} \tanh \frac{\pi}{\beta} t, \qquad E = \frac{\pi \phi_r}{4\beta^2 G}. \tag{2.4.1}$$

The metric and dilaton profile are [50]

$$\mathrm{d}s^2 = -\frac{4\pi^2}{\beta^2} \frac{\mathrm{d}u \mathrm{d}v}{\sinh^2 \frac{\pi}{\beta}(u-v)}, \qquad \phi = \frac{2\pi\phi_r}{\beta} \frac{1}{\tanh \frac{\pi}{\beta}(u-v)}. \tag{2.4.2}$$

For $E \geq 0$ this solution corresponds to a black hole with inverse temperature β. The identification of β with the inverse temperature can be seen by requiring the Euclidean section of the spacetime to be regular at the horizon, as discussed at the end of section 2.1.1. This suggests that the vacuum state should look thermal to an asymptotic observer. Since physics happens in real time, a more satisfying argument is the one given in section (2.2) which showed that the conformal anomaly together with the condition of a regular state at the horizon imply that there must be a flux of thermal radiation at late times. Since this argument was quite general, it's instructive to see how this works for a real, massless scalar field φ. The expectation value of the right-moving component of the stress tensor

$$T_{uu} = \partial_u \varphi \, \partial_u \varphi, \tag{2.4.3}$$

can be extracted from the two-point function of φ. Expanding φ in modes $f_\omega, \overline{f}_\omega$ of positive and negative frequency with respect to the Poincaré vacuum

$$\varphi(u,v) = \int_0^\infty \mathrm{d}\omega \left(a_\omega f_\omega + a_\omega^\dagger \overline{f}_\omega \right), \qquad f_\omega = \frac{e^{-\mathrm{i}\omega F(u)} - e^{-\mathrm{i}\omega F(v)}}{\sqrt{4\pi\omega}}, \tag{2.4.4}$$

such that the Dirichlet boundary condition $\varphi = 0$ is satisfied at the boundary $u = v$, we find

$$\langle\varphi(u,v)\varphi(u',v')\rangle = -\frac{1}{4\pi}\log\left|\frac{(F(u)-F(u'))(F(v)-F(v'))}{(F(u)-F(v'))(F(v)-F(u'))}\right|. \tag{2.4.5}$$

This just usual two-point function of a real, massless scalar field with an image term which gives rise to the denominator. Using the explicit form of the function F (2.4.1) the correlation function reads

$$\langle\varphi(u,v)\varphi(u',v')\rangle = -\frac{1}{4\pi}\log\left|\frac{\sinh\frac{\pi}{\beta}(u-u')\sinh\frac{\pi}{\beta}(v-v')}{\sinh\frac{\pi}{\beta}(u-v')\sinh\frac{\pi}{\beta}(v-u')}\right|. \tag{2.4.6}$$

Notice that this is periodic in imaginary time with period β. This is the KMS condition [53, 54] that must hold for any thermal correlation function and follows simply from cyclicity of the trace and that the logarithm of the thermal density matrix is the Hamiltonian.

Classically, for the scalar field φ the right-moving component of the stress tensor is given by (2.4.3), but as a composite operator in the quantum theory this operator requires renormalisation due to the universal short distance behaviour in the operator product expansion

$$\partial_u\varphi\,\partial_{u'}\varphi \overset{u\to u'}{=} -\frac{1}{4\pi}\frac{1}{(u-u')^2} + \text{reg}, \tag{2.4.7}$$

where reg denotes terms which are regular as $u \to u'$. The divergence is universal in the sense that it is present for any state. The renormalised or normal ordered stress tensor $:T_{ab}:$ in the quantum theory can be defined by point-splitting. For the right-moving component this is defined as

$$:T_{uu}: = \lim_{u'\to u}\left[\partial_u\varphi\,\partial_{u'}\varphi + \frac{1}{4\pi}\frac{1}{(u-u')^2}\right]. \tag{2.4.8}$$

There is a similar expression for the left-moving component. Notice that this definition refers to a specific frame as it is defined by subtracting the singular part of the composite operator $\partial_u\varphi\,\partial_{u'}\varphi$ in the frame associated to an asymptotic observer, that is, an observer who defines positive frequency with respect to boundary time t. Using (2.4.5) and the explicit form of the function F (2.4.1) it is straightforward to compute

$$\langle:T_{uu}:\rangle = -\frac{1}{24\pi}\{F,u\} = \frac{\pi}{12\beta^2}, \tag{2.4.9}$$

which implies thermal flux of radiation at inverse temperature β. This is precisely (2.2.5) with $c = 1$ for real, massless scalar field. Notice that the normal ordered stress tensor transforms inhomogeneously.

Since we have identified β as the inverse temperature of the black hole we can associate a classical thermodynamic entropy to the black hole by integrating the first law of thermodynamics

$$S_{\rm BH} = \int \mathrm{d}E\, \beta(E) = S_0 + \frac{2\pi\phi_r/\beta}{4G}, \tag{2.4.10}$$

with $S_0 = \phi_0/4G$. This is just $A/4G$ with A the value of the dilaton $\phi_0 + \phi$ on the horizon. Notice that $S_{\rm BH} - S_0$ is linear in the temperature, as expected for a near-extremal black hole in higher dimensions.

Although we will primarily be interested in nonequilibrium solutions, which correspond to evaporating black holes, we will find that over short time scales compared with the time scale for evaporation the solution is well approximated by the equilibrium solution. This follows since black holes evaporate very slowly in the semiclassical limit so are effectively in equilibrium with their Hawking radiation over sufficiently short time scales.

2.5 Evaporating black holes

To model evaporation we couple the matter CFT to an auxiliary reservoir system which is taken to be the same matter CFT defined on half of Minkowski space with metric $\mathrm{d}s^2 = -\frac{1}{\epsilon^2}\mathrm{d}u\mathrm{d}v$. In coupling these systems, we identify the boundary time with Minkowski time and impose transparent boundary conditions for the matter fields so that energy and momentum flows freely across the interface. We can think of this as a toy model describing the s-wave sector of a near-extremal black hole in higher dimensions but where the near-horizon region of the black hole (sometimes called the "zone"), which corresponds to the nearly AdS_2 factor, has been glued to the asymptotic region, which corresponds to the half Minkowski space.

In this setup we can form an evaporating black hole by heating up a zero-temperature black hole and letting it evaporate into the reservoir [55]. In more detail, we start with the extremal black hole in JT gravity with matter in the Hartle-Hawking state (or Poincaré vacuum). At time $t = 0$, we throw in a pulse of energy from the right boundary. Such a pulse of energy can be created by a "local quench" in the matter CFT, created by inserting a scalar primary operator with conformal dimension Δ at a point on the boundary at $t = \mathrm{i}\varepsilon$. The shift in imaginary time by parameter ε acts a regulator, ensuring that the state has finite norm. The parameter ε should not be confused with the parameter ϵ which enters in the boundary conditions for JT gravity. The expectation value of the stress tensor in this state is determined by the conformal Ward identity,

$$\langle :T_{vv}: \rangle = \frac{\Delta}{\pi} \frac{\varepsilon^2}{(v^2+\varepsilon^2)^2} \to E_0\, \delta(v). \tag{2.5.1}$$

The RHS holds in the limit $\varepsilon \to 0$ with $E_0 = \Delta/2\varepsilon$ held fixed and describes a shockwave incident on the black hole (see figure 2.2). As shown in section 2.2 any sensible state for the outgoing modes will be extremely well approximated by the Kruskal vacuum for which $\langle T_{UU} \rangle = 0$.

2.5.1 Solving for F

In the presence of a local quench the energy balance equation (2.3.7) reads

$$\frac{\mathrm{d}E}{\mathrm{d}t} = E_0 \delta(t) - kE, \qquad k = \frac{cG}{3\phi_r}. \tag{2.5.2}$$

The parameter $1/k$ will turn out to set the time scale for evaporation. To the past of the shockwave we have the extremal black hole solution which has $F(t) = t$ and $E = 0$. To the future of the shockwave F solves

$$\{F, t\} = -\frac{8\pi G}{\phi_r} E_0 e^{-kt}, \qquad t < 0. \tag{2.5.3}$$

By considering the associated linear second order ODE (2.3.9) one finds an exact solution in terms of modified Bessel functions

$$\widetilde{F}(t) = \frac{e^{2z_0}}{\pi} \frac{K_0(z)}{I_0(z)}, \qquad z = \frac{2\pi T_0}{k} e^{-kt/2}, \tag{2.5.4}$$

where $z_0 = z(t = 0)$. The parameter T_0 defined through the relation

$$E_0 = \frac{\pi \phi_r}{4G} T_0^2, \tag{2.5.5}$$

will later be shown to represent the initial temperature of the heated up black hole. In the presence of a delta function shockwave F cannot be a smooth function. The best we can do is ensure it is continuous up to its second derivative. Writing

$$F = \frac{\mathsf{a}\widetilde{F} + \mathsf{b}}{\mathsf{c}\widetilde{F} + \mathsf{d}}, \tag{2.5.6}$$

and imposing continuity $F(0) = 0, F'(0) = 1$, and $F''(0) = 0$ fixes

$$\mathsf{a} = \frac{2\pi}{kz_0} e^{-2z_0} I_0(z_0)\,, \quad \mathsf{b} = -\frac{2}{kz_0} K_0(z_0)\,, \quad \mathsf{c} = \pi e^{-2z_0} I_1(z_0)\,, \quad \mathsf{d} = K_1(z_0), \tag{2.5.7}$$

up to an irrelevant positive scalar, and hence determines F in terms of $\widetilde{F}$.

In the semiclassical limit the time scale for evaporation $1/k$ is very large and there is simple yet excellent approximation to the exact solution (2.5.4) which is valid until very late times. The approximate formula can be derived from the asymptotic formula

for the modified Bessel functions. A more intuitive approach is to note that the linear second order ODE associated to (2.5.3) is of WKB form

$$\varepsilon^2 f'' = e^{-x} f, \tag{2.5.8}$$

with dimensionless parameters $\varepsilon = k/\pi T_0$, $x = kt$, and $' = \frac{\mathrm{d}}{\mathrm{d}x}$. For $\varepsilon \ll 1$ with $x = kt$ fixed we find two solutions

$$f_\pm(t) \sim e^{x/4} \exp\left(\pm \frac{1}{\varepsilon} \int_0^x \mathrm{d}x' e^{-x'/2}\right). \tag{2.5.9}$$

Taking the ratio gives

$$\widetilde{F}(t) \sim \exp\left(2\pi \int_0^t \mathrm{d}t'\, T(t')\right), \qquad T(t) = T_0 e^{-kt/2}. \tag{2.5.10}$$

The function $T(t)$ will later be shown to represent the instantaneous temperature of the black hole. The approximate solution for F is found by applying a Möbius transformation

$$F(t) \sim \frac{1}{\pi T_0} \tanh\left(\pi \int_0^t \mathrm{d}t'\, T(t')\right). \tag{2.5.11}$$

Notice that over time scales $t \ll 1/k$ the temperature is effectively constant and so the solution is the approximately the equilibrium solution (2.4.1). We can understand when the approximation breaks down by going to higher order in the WKB expansion. Including the first correction gives

$$\log \widetilde{F}(t) \sim \frac{4\pi T_0}{k}\left(1 - e^{-kt/2}\right) + \frac{k}{8\pi T_0}\left(1 - e^{kt/2}\right). \tag{2.5.12}$$

For the RHS to be an asymptotic series in k/T_0 we require the first term to be much larger than second term. This indeed holds for $k/T_0 \ll 1$ with kt held fixed. However, for late times $t \sim \frac{1}{k} \log \frac{T_0}{k}$ the second term becomes larger which signifies a breakdown of the WKB approximation.

Indeed, for late times the black hole eventually cools back down to zero temperature. This follows simply because for late times (2.5.8) is approximately solved by a function which is linear in time, which corresponds to the solution for the extremal black hole. The transition between these regimes can be understood from the exact solution, given in terms of a ratio of modified Bessel functions (2.5.4). For example, the exact solution determines the location of the horizon. In Kruskal coordinates,

$$U_{\text{horizon}} = -\lim_{u \to \infty} 1/\widetilde{F}(u) = 0. \tag{2.5.13}$$

The late time behaviour will not be important for us. For instance, the Page time occurs when $S_{\text{BH}} - S_0$ is an order one fraction of its initial size, which happens at a time scale of order $1/G$. More specifically, in JT gravity the Page time occurs at a time scale of order $1/k$ which is well within the regime described by the approximate solution (2.5.10).

2.5.2 Entropy of Hawking radiation

In Kruskal coordinates, the metric and dilaton profile for the evaporating black hole are [50, 55]

$$\mathrm{d}s^2 = -\frac{4\mathrm{d}U\mathrm{d}V}{(1+UV)^2}, \qquad \phi = \phi_r\left(\frac{\widetilde{F}''(v)}{\widetilde{F}'(v)} - \frac{1}{2}\frac{U\widetilde{F}'(v)}{1+UV}\right). \tag{2.5.14}$$

The Kruskal coordinates are related to the Minkowski lightcone coordinates u and v, adapted to an asymptotic observer, by

$$U = -1/\widetilde{F}(u), \qquad V = \widetilde{F}(v), \tag{2.5.15}$$

where the solution for $\widetilde{F}$ is given in (2.5.10). Notice that this looks like a simple generalisation of (2.1.15) which accounts for the fact that the temperature of the black hole varies slowly with time. It follows from the argument in section (2.2) that there is an outgoing flux of thermal Hawking radiation

$$\langle{:}T_{uu}{:}\rangle = \frac{c}{24\pi}\{U,u\} \approx \frac{\pi c}{12}T(u)^2. \tag{2.5.16}$$

By integrating the first law we can associate a classical thermodynamic entropy to the black hole

$$S_{\mathrm{BH}}(v) \approx S_0 + \frac{2\pi\phi_r T(v)}{4G}. \tag{2.5.17}$$

This is a simple generalisation of the formula (2.4.10) for the eternal black hole which accounts for the fact that for an evaporating black hole the temperature is slowly varying with time. These formulas also confirm the fact that over short time scales compared with the time scale $1/k$ for evaporation, where the temperature of the black hole is effectively constant, the solution is well approximated by the equilibrium solution.

The Hawking radiation also has an entropy. The entropy of all the radiation emitted since the formation of the black hole at $t = 0$ up to some time t, taken to be much larger than the thermal scale, is given by the thermodynamic formula

$$S_{\mathrm{rad}}(t) = \frac{\pi c}{6}\int_0^t T(t')\mathrm{d}t'. \tag{2.5.18}$$

The entropy grows monotonically with time and eventually we run into a puzzle when it becomes larger than the Bekenstein-Hawking entropy S_{BH}. As discussed in the introduction, the entropy cannot continue to grow forever. By the Page time, it must turn over and eventually decrease to zero. As we will see in the next section, a resolution to this puzzle is provided by the QES prescription, which shows that the entropy of the Hawking radiation is only given by (2.5.18) before the Page time. The Page time is the point at which $S_{\mathrm{rad}} = S_{\mathrm{BH}}$. After the Page time, the entropy of the radiation begins to decrease. In particular, there does not appear to be any obvious obstruction to the entropy not eventually decreasing to zero at the endpoint of evaporation. However, near the endpoint of evaporation, we are well outside the semiclassical regime.

2.5.3 Page curve

To compute the entropy all the radiation emitted since the formation of the black hole up to some time t we consider an interval in the reservoir with one endpoint on the boundary $u = v = t$ and one endpoint at spatial infinity. We will consider the possibility of a nontrivial QES which lies somewhere in the AdS_2 region. The generalised entropy is

$$S_{\text{gen}}(I \cup R) = \frac{A(\partial I)}{4G} + S_{\text{QFT}}(I \cup R), \tag{2.5.19}$$

where I is a spatial interval (or more precisely the domain of dependence of the interval), called the *island*, which ends on the QES and extends to spatial infinity in the other direction. The term *island* arises because in certain cases I is spatially disconnected from R. A more fitting name in this context might be *peninsula* as I extends out from the other spatial infinity [56].

In JT gravity the area term corresponds to the dilaton, that is, $A = \phi_0 + \phi$. The second term is the entropy of quantum fields in $I \cup R$ in the Unruh state. For a CFT the formula for this entropy is universal and only depends on the central charge c of the CFT

$$S_{\text{QFT}}(I \cup R) = \frac{c}{6} \log \frac{-(U - U_{\partial R})(v - v_{\partial R})}{\Omega_{\partial I}\Omega_{\partial R}}, \tag{2.5.20}$$

where U and v are coordinates of the QES and $U_{\partial R} = -1/\widetilde{F}(t)$ and $v_{\partial R} = t$ are the coordinates of ∂R. In this expression $\Omega_{\partial I}$ and $\Omega_{\partial R}$ are Weyl factors, associated to the endpoints ∂I and ∂R, which arise from rescaling the AdS_2 and Minkowski metric, respectively, to the metric $-\mathrm{d}U\mathrm{d}v$. Explicitly,

$$\Omega_{\partial I}^{-2} = \frac{2\pi T(v)V}{(1+UV)^2}, \qquad \Omega_{\partial R}^{-2} = -\frac{1}{2\pi T(u)U}. \tag{2.5.21}$$

The contribution from the Weyl factors in (2.5.20) can be understood as arising from the Weyl anomaly, which constrains the difference between the entropy computed in two different frames to be

$$S_{\Omega^2 g} - S_g = \frac{c}{6} \sum_{\text{endpoints}} \log \Omega, \tag{2.5.22}$$

where the sum is over the endpoints of the interval that we are computing the entropy of. This follows from the fact that the entropy in a CFT is related to a correlation function of primary fields called twist and antitwist fields which have conformal weights proportional to the central charge c, as we explain in section 5.3.1.

The location of the QES is determined by extremising the generalised entropy. Assuming the QES lies close to the horizon, we can approximate the area term as

$$\frac{A}{4G} \approx S_0 + (S_{\text{BH}}(v) - S_0)\,(1 - 2UV). \tag{2.5.23}$$

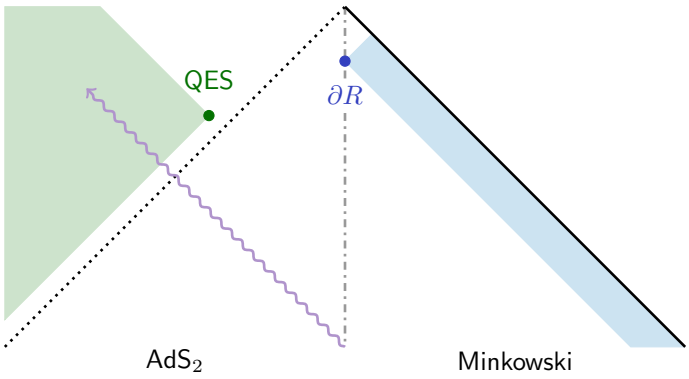


Figure 2.2: The entanglement wedge of the radiation of an old evaporating black hole contains a portion of the interior (shaded green) of the black hole. The wiggly lilac line represents the ingoing shockwave which formed the black hole.

Extremising the generalised entropy with respect to v then gives

$$\frac{\partial S_{\text{gen}}}{\partial v} \approx -4\pi T(v)\left(S_{\text{BH}}(v) - S_0\right) UV + \partial_v S_{\text{BH}}(v) + \frac{\pi c}{6} T(v). \tag{2.5.24}$$

The negative flux of energy across the horizon together with the first law imply the middle term is equal to

$$\partial_v S_{\text{BH}}(v) = -\frac{\pi c}{12} T(v). \tag{2.5.25}$$

This leads to

$$UV \approx \frac{c}{48\left(S_{\text{BH}}(v) - S_0\right)}. \tag{2.5.26}$$

In the semiclassical limit the RHS is very small and so the QES lie inside and very near to the horizon. Extremising the generalised entropy with respect to U gives

$$\frac{\partial S_{\text{gen}}}{\partial U} \approx -2(S_{\text{BH}}(v) - S_0)V + \frac{c}{6}\frac{1}{U - U_{\partial R}}. \tag{2.5.27}$$

Using (2.5.26), the solution is

$$U = -\frac{1}{3U_{\partial R}}. \tag{2.5.28}$$

This means that, to leading order, the location of the QES is

$$\tilde{u} = t, \qquad v = t - \frac{1}{2\pi T(t)} \log \frac{S_{\text{BH}}(t) - S_0}{c}, \tag{2.5.29}$$

where we have dropped correction of order the thermal scale. Here $\tilde{u}$ is defined in a similar way to (2.1.16) by the relation $U = 1/\widetilde{F}(\tilde{u})$. Since the QES lies near the horizon, the area term is approximately given by the Bekenstein-Hawking entropy evaluated at time t, that is, $A/4G \approx S_{\text{BH}}(t)$. Also, since the island lies behind the horizon, it

contains the modes which are entangled with the outgoing Hawking radiation in the Unruh state (2.2.6) and so $S_{\text{QFT}}(I \cup R) \approx 0$. This will be explained in more detail in the next section. Hence, the generalised entropy is

$$S_{\text{gen}}(I \cup R) \approx S_{\text{BH}}(t). \tag{2.5.30}$$

From the QES prescription, the entropy of the radiation subsystem R is given by

$$S(R) \approx \min \left\{ S_{\text{rad}}(t), S_{\text{BH}}(t) \right\}, \tag{2.5.31}$$

which is the Page curve. Notice that the ingoing coordinate v (2.5.29) of the QES lags behind by a scrambling time

$$t_{\text{scr}} = \frac{1}{2\pi T(t)} \log \frac{S_{\text{BH}}(t) - S_0}{c}. \tag{2.5.32}$$

This means that any information thrown into the black hole after the Page time and at least a scrambling time in the past will lie in the island. According to entanglement wedge reconstruction, this information can be recovered from the radiation after only a scrambling time, as predicted by Hayden and Preskill [47]. In section 4 we will investigate aspects of this problem in more detail.

There is a subtlety that we have brushed over here. The entropy $S_{\text{rad}}(t)$ does not arise from an empty QES, that is, it is not true that $S_{\text{QFT}}(R) \approx S_{\text{rad}}(t)$. Instead, $S_{\text{QFT}}(R)$ is infrared divergent. This itself is a version of the information problem. It turns out that the resolution to this puzzle is also to include a nontrivial QES [55, 56]. This QES has generalised entropy

$$S_{\text{gen}}(I_0 \cup R) \approx S_0 + S_{\text{rad}}(t). \tag{2.5.33}$$

So the entropy as stated in (2.5.31) is really only correct if the extremal entropy is small, otherwise we should add the extremal entropy to the term which contributes as $S_{\text{rad}}(t)$. An alternative way to get around the issue is to cutoff the infrared divergence by not taking R to extend all the way to spatial infinity.

Chapter 3

Islands, irreversibility, and greybody factors

This chapter primarily consists of two papers [57, 58] written in collaboration with Timothy J. Hollowood, S. Prem Kumar, and Andrea Legramandi. Some results in section 3.7.4 rely on appendix B.3 which is adapted from a paper [59] coauthored additionally with Zsolt Gyongyosi.

3.1 Introduction

Inspired by earlier arguments by Bekenstein [2], Hawking famously showed that black holes are not completely black but instead have a temperature [3]. By the first law of thermodynamics one can then associate a classical thermodynamic entropy to the black hole

$$S_{\rm BH} = \frac{A}{4G}, \tag{3.1.1}$$

where A is the horizon area of the black hole. Hawking's calculation showed that, from the perspective of perturbative semiclassical gravity, the radiation emitted by a black hole is in a mixed thermal state (up to greybody factors) at the Hawking temperature, regardless of how the black hole was formed. This leads to an apparent tension with unitarity if the radiation continues to remain in a mixed state even after the black hole has entirely evaporated [4]. This traditional version of the information problem is naive since it relies on extrapolating Hawking's semiclassical calculation to arbitrarily late times, when the black hole has almost entirely evaporated.

A sharper formulation of the information problem hinges on viewing the black hole, B, and its radiation, R, as a bipartite quantum system and the following bound on the

von Neumann entropy $S(R)$ of the radiation subsystem

$$S(R) \leq \min\{S_{\rm rad}, S_{\rm BH}\}, \tag{3.1.2}$$

which follows if one believes that the entropy of the black hole $S_{\rm BH}$ has a statistical mechanical origin. In this expression $S_{\rm rad}$ is the thermal entropy of the radiation which follows from Hawking's calculation. Since the thermal entropy $S_{\rm rad}$ increases monotonically with time we run into a contradiction when it becomes larger than $S_{\rm BH}$. The time at which this occurs is known as the Page time. This sharpens the original argument because at the Page time, which requires an $\mathcal{O}(1/G)$ time evolution, the black hole is still an $\mathcal{O}(1)$ fraction of its initial size. Actually, by modelling the microscopic dynamics of the black hole as a random unitary transformation, Don Page [5] went further and argued that the upper bound is approximately saturated

$$S(R) \approx \min\{S_{\rm rad}, S_{\rm BH}\}. \tag{3.1.3}$$

This result is known as the Page curve. Recent work [9–12] has shown how to obtain the Page curve using the quantum extremal surface (QES) prescription, as was briefly reviewed in section 2.5.31. The QES (or island) prescription says that the entropy of the radiation R is given by[1]

$$S(R) = \min_{I} \operatorname{ext} S_{\rm gen}(I \cup R) = \min_{I} \operatorname{ext} \left[\frac{A(\partial I)}{4G} + S_{\rm QFT}(I \cup R) \right]. \tag{3.1.4}$$

The generalised entropy $S_{\rm gen}(I \cup R)$ splits into a piece proportional to the area $A(\partial I)$ of the codimension two surface ∂I (known as the QES) of the region I (known as the island) and the von Neumann entropy $S_{\rm QFT}(I \cup R)$ of quantum fields in $I \cup R$ as computed in pure QFT with no gravity. Hawking's calculation corresponds to only considering the empty QES and contributes as $S_{\rm QFT}(R) \approx S_{\rm rad}$. In two dimensions the QES ∂I is a point or a collection of points and in JT gravity the area A is replaced by the dilaton $\phi_0 + \phi$.

The QES prescription also applies when R is any subset of the radiation, however, in general it's a daunting task to analytically determine the extrema of the generalised entropy for an arbitrary subset. The first aim of this work is to show how this problem greatly simplifies in the semiclassical limit where black holes evaporate very slowly. In this "adiabatic" limit it is meaningful to associate a slowly varying temperature $T(t)$ to the black hole and the entropy carried away by a subset $R = \bigsqcup_k R_k$ of the radiation is simply given by the thermodynamic formula

$$S_{\rm QFT}(R) \approx S_{\rm rad}(R), \qquad S_{\rm rad}(R) = \frac{\pi N}{6} \int_R T(u) \mathrm{d}u, \tag{3.1.5}$$

[1] Various aspects of islands have been explored in [56, 60–71]. For an alternative perspective, a discussion on the subtleties of separating the black hole and radiation subsystems, and critiques, see [72–78].

where N counts the number of massless fields. The thermodynamic formula (3.1.5) for the entropy applies when the subsets of R and the spaces between them are large compared with the thermal scale, that is, measured with respect to $T(u)\mathrm{d}u$ or $\frac{\mathrm{d}U}{|U|}$. Throughout this chapter we assume that R is of this form and simply refer to this assumption as the thermodynamic limit. In (3.1.5) the integral over R denotes an integral over the range of the outgoing null coordinates which span R. That is, if the range of outgoing null coordinates u over the interval R_k is $[u_k, u'_k]$, then

$$S_{\mathrm{rad}}(R_k) = \frac{\pi N}{6} \int_{u_k}^{u'_k} T(u)\mathrm{d}u. \tag{3.1.6}$$

By combining the adiabatic and thermodynamic limit with the QES formula we find a very simple recipe for determining a class of islands I which can contribute to the entropy of a subset R of the radiation:

1. Choose a subset $\partial\tilde{I} \subset \partial R$. Here $\tilde{I}$ (the "island in the stream") is the image of the island I under the involution $U \to -U$, where U denotes the outgoing null Kruskal coordinate, which exchanges the interior and exterior region of the black hole.

2. The generalised entropy of $I \cup R$ is given by

$$S_{\mathrm{gen}}(I \cup R) = \sum_{\substack{\text{disconnected} \\ \text{components} \\ \partial\tilde{I}_k \text{ of } \partial\tilde{I}}} S_{\mathrm{BH}}(u_{\partial\tilde{I}_k}) + S_{\mathrm{rad}}(\tilde{I} \ominus R), \tag{3.1.7}$$

   where $\ominus$ denotes the symmetric difference, $\tilde{I} \ominus R = \tilde{I} \cup R - \tilde{I} \cap R$. Finally the von Neumann entropy $S(R) = \min_I S_{\mathrm{gen}}(I \cup R)$ of R is determined by minimising over possible islands.

A few clarifications are necessary. First, although we have stated the formula in general terms, its derivation is based on a simple model of JT gravity coupled to an auxiliary reservoir, with the matter sector consisting of N massless Dirac fermions.[2] The choice of matter sector is dictated by the fact that Dirac fermions are the only CFT for which the multi-interval von Neumann entropy is known analytically [79]. Second, we cannot exclude the possibility of additional islands which could have lower generalised entropy and hence contribute to the entropy of R. However, in section 3.4 we verify that various entropy inequalities, including strong subadditivity, are obeyed providing a nontrivial consistency check on the class of islands we find.

The island in the stream recipe, as described above, is only valid under the simplifying assumption of a trivial greybody factor. A feature of a black holes in higher

[2]Each Dirac fermion has central charge 1 so the total central charge is $c = N$.

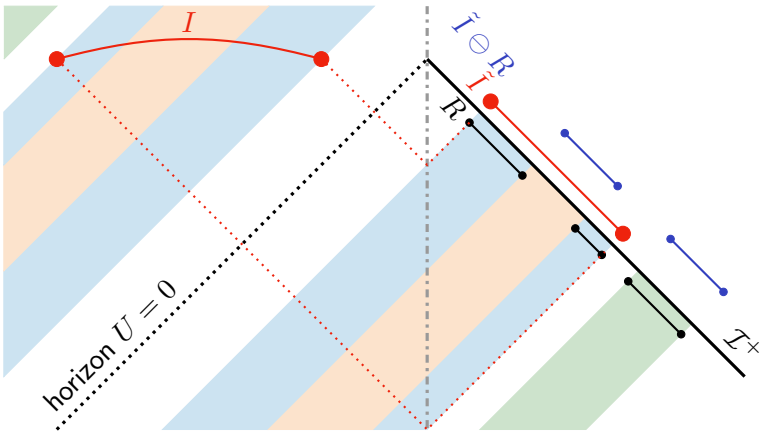


Figure 3.1: An example where R consists of three intervals with an island I behind the horizon shown in red and its reflection across the horizon $\tilde{I}$ also shown in red. The coloured bands represent different subsets of outgoing modes and their entangled partner modes behind the horizon in the same colour. The blue modes do not contribute to $S_{\mathsf{QFT}}(I \cup R)$ since both the outgoing modes and their entangled partners are contained in $I \cup R$. On the other hand, the green and orange modes do contribute since only one of the outgoing modes or their partners are contained in $I \cup R$.

dimensions, as opposed to the simple model of JT gravity, is the fact that Hawking radiation must tunnel through an effective potential barrier around the black hole. This means that only some of the radiation escapes out to infinity and the spectrum of Hawking radiation gets modified from a blackbody to a greybody [80]

$$\langle N_\omega^{\mathbb{T}} \rangle = \frac{\Gamma(\omega)}{e^{\omega/T} - 1}. \tag{3.1.8}$$

The greybody factor $\Gamma(\omega)$ is the transmission probability for a mode with definite frequency ω to tunnel through the potential barrier and $\langle N_\omega^{\mathbb{T}} \rangle$ denotes the expectation value of the number operator associated to the transmitted mode $\mathbb{T}$. For certain black holes, like Schwarzschild, the greybody factor leads to simple multiplicative relation between the entropy flux $\mathrm{d}S_{\mathrm{rad}}^{\mathbb{T}}/\mathrm{d}u$ of the transmitted radiation and the entropy flux of the Bekenstein-Hawking entropy

$$\frac{\mathrm{d}S_{\mathrm{rad}}^{\mathbb{T}}}{\mathrm{d}u} = -\xi \frac{\mathrm{d}S_{\mathrm{BH}}}{\mathrm{d}u} \qquad \Rightarrow \qquad \frac{\mathrm{d}}{\mathrm{d}u}\left(S_{\mathrm{BH}} + S_{\mathrm{rad}}^{\mathbb{T}}\right) = (\xi - 1)\left|\frac{\mathrm{d}S_{\mathrm{BH}}}{\mathrm{d}u}\right|. \tag{3.1.9}$$

Here the parameter $1 \leq \xi \leq 2$ is the "greybody coefficient". For example, for Schwarzschild $\xi = 1.91$ in the s-wave approximation. The limit $\xi \to 1$ is the reversible limit where all the radiation is reflected back into the black hole while $\xi = 2$ is the case with a trivial greybody factor $\Gamma = 1$ where all the radiation escapes out to infinity. Note that $S_{\mathrm{BH}} + S_{\mathrm{rad}}^{\mathbb{T}}$ is nondecreasing since $\xi \geq 1$. This is the statement of the generalised second law [2, 81].

The aim of section 3.6 is to understand the effect of a greybody factor on the entropy of a subset R of radiation, and in particular how it affects the Page curve and the structure of correlations in the radiation. Working in the adiabatic and thermodynamic limit[3] we find that the formula for the entropy given by a simple extension of the island in the stream formula (3.1.7)

$$S_{\text{gen}}(I \cup R) = \sum_{\substack{\text{disconnected components} \\ \partial\tilde{I}_k \text{ of } \partial\tilde{I}}} S_{\text{BH}}(u_{\partial\tilde{I}_k}) + S^{\mathbb{T}}_{\text{rad}}(\tilde{I} \ominus R). \tag{3.1.10}$$

As compared with (3.1.7) the second term now involves the entropy $S^{\mathbb{T}}_{\text{rad}}$ of the radiation which is transmitted through the potential barrier and has a deformed Planck spectrum (3.1.8). The greybody factors are of higher dimensional black holes are absent in the simple model of JT gravity. However, we mimic the effect of a greybody factor by picking certain boundary conditions at the AdS-Minkowski reservoir interface so that modes are only partly transmitted.

A feature which becomes apparent from the islands in the stream formula is that in contrast to the Page curve calculation where one considers the entropy of all the radiation, for subsets of the radiation there can be a nontrivial competition between multiple islands. Naively this is at odds with Page's calculation which says that the entropy of R should be given by minimising the thermodynamic entropies of the subsystem R and the complementary subsystem $B \cup R'$, where R' is the complement of R in the radiation, that is,

$$S(R) \overset{?}{\approx} \min\{S^{\mathbb{T}}_{\text{rad}}(R), S_{\text{BH}} + S^{\mathbb{T}}_{\text{rad}}(R')\}. \tag{3.1.11}$$

The resolution to this potential puzzle is that Page's toy model assumes the process of black hole evaporation is thermodynamically reversible. In section 3.7 we consider a generalisation of Page's model [82] which accounts for irreversibility and demonstrate how the entropies perfectly match with the island in the stream formula (3.1.10). In particular, viewing the greybody coefficient ξ as a free parameter allows us to investigate the effect of irreversibility on the entropy of the radiation. A key result is that irreversibility of evaporation leads to the possibility of multiple islands which can contribute to the entropy of a generic subset R of radiation. This is in sharp contrast to the reversible limit $\xi \to 1$ where only one nonempty island can contribute.

The plan of this chapter is as follows. We start in section 3.2 by reviewing some aspects of evaporating black holes in JT gravity and the entanglement structure of quantum fields across the horizon. In section 3.3 we derive the islands in the stream

[3] Actually we require a more stringent limit where the subsets of radiation and spaces between them are not only much larger than the thermal scale but also much larger than the scrambling time $t_{\text{scr}} = \frac{1}{2\pi T} \log \frac{S_{\text{BH}} - S_0}{N}$. This additional requirement is not entirely necessary but relaxing it leads to a more complicated formula for the entropy.

formula (3.1.7), deferring some technical details to appendix B.1. In section 3.4 we discuss some applications of the islands in the stream formula and show that various constraints from unitarity and entropy inequalities are obeyed. The role of greybody factors on evaporation and entropies is discussed in section 3.5 and the generalisation (3.1.10) of the island in the stream formula to account for greybody factors is derived in section 3.6. In section 3.7 we review Page's model before discussing a generalisation of the model which accounts for irreversibility. We investigate the role of irreversibility on the entropy and compare the results with the island in the stream formula. Finally, in section 3.8 we end with some conclusions.

3.2 Evaporating black holes and entanglement

In this section we review some aspects of evaporating black holes in JT gravity and the entanglement structure of quantum fields across the horizon.

3.2.1 Evaporating black holes

JT gravity captures the dynamics of the two-dimensional effective theory governing the s-wave sector of the near-horizon limit of near-extremal black holes in higher dimensions. To study evaporating black holes, JT gravity can be coupled to an auxiliary reservoir. One scenario involves heating up a zero-temperature black hole and letting it evaporate into the reservoir, as described in section 2.5. However, the specific details of the black hole's formation will not be relevant for this work.

The spacetime in JT gravity is fixed to be a portion of AdS_2, which we can parameterise in Kruskal coordinates

$$\mathrm{d}s^2 = -\frac{4\mathrm{d}U\mathrm{d}V}{(1+UV)^2}. \tag{3.2.1}$$

The future event horizon is the surface $U = 0$. Near the conformal boundary $UV = -1$ the geometry is glued to an auxiliary reservoir which is taken to be a copy of half of Minkowski space. These coordinates are related to the Minkowski lightcone coordinates u and v, adapted to an asymptotic observer, by[4]

$$U = -1/F(u), \qquad V = F(v). \tag{3.2.2}$$

The dynamics of JT gravity is encoded a boundary action for the function F which relates the Kruskal and Minkowski frames. In particular, the dilaton is determined in terms of this function

$$\phi = \phi_r \left(\frac{F''(v)}{F'(v)} - \frac{1}{2}\frac{UF'(v)}{1+UV} \right). \tag{3.2.3}$$

[4]In chapter 2, the functions F and $\widetilde{F}$ related the Poincaré and Kruskal frames, respectively, to the frame of an asymptotic observer. But since the Poincare frame won't play a role in any of the remaining chapters we will instead write F for $\widetilde{F}$, that is, $F_{\text{chapters 3,4,5}} = \widetilde{F}_{\text{chapter 2}}$.

The exact solution for the function F can be expressed in terms of modified Bessel functions (2.5.4), however, there is an simple yet excellent WKB like solution (2.5.10) which is valid until very late times,

$$F(t) \approx \exp\left(2\pi \int_0^t \mathrm{d}t'\, T(t')\right), \qquad T(t) = T_0 e^{-kt/2}. \tag{3.2.4}$$

In the semiclassical limit $k = cG/3\phi_r \ll 1$ and the slowly varying function T can be interpreted as the instantaneous temperature of the black hole. Notice that the relation between the Kruskal and Minkowski frame is a simple generalisation of usual relation (2.1.15) between the coordinates for an eternal black hole.

3.2.2 The quantum state

For an evaporating black hole the relevant state for quantum fields is the Unruh state. This is the state where the outgoing modes are in the vacuum associated to the Kruskal frame and the ingoing modes are in the vacuum associated to the frame of an asymptotic observer. Hawking showed that the spectrum of the radiation which tunnels through the effective potential barrier around the black hole in the Unruh state is determined by

$$\langle N_\omega^{\mathbb{T}} \rangle = \frac{\Gamma(\omega)}{e^{\omega/T} - 1}. \tag{3.2.5}$$

In this expression $\langle N_\omega^{\mathbb{T}} \rangle$ denotes the expectation value of the number operator associated to the transmitted mode $\mathbb{T}$ with definite frequency ω and the greybody factor $\Gamma(\omega)$ is the transmission probability for a mode to tunnel through the potential barrier. For simplicity, we will initially ignore mixing of ingoing and outgoing modes caused by the potential barrier and set $\Gamma(\omega) = 1$. With this simplification the spectrum of Hawking radiation is thermal. This can also be seen from the conformal anomaly. Since the outgoing modes are in the vacuum associated to the Kruskal frame, we find

$$\langle :T_{uu}: \rangle = \frac{c}{24\pi}\{U, u\} \approx \frac{\pi c}{12} T(u)^2. \tag{3.2.6}$$

This is nothing but an outgoing flux of thermal Hawking radiation. In JT gravity, ignoring the mixing between the ingoing and outgoing modes caused by the potential barrier amounts to choosing transparent boundary conditions for quantum fields along the AdS-Minkowski interface. We will address the mixing of the modes and a nontrivial greybody in section 3.5.

3.2.3 Thermal entropy

During an outgoing null time δu which is small, but much larger than the thermal scale, the black hole emits thermal radiation which fills a shell of thickness δu, in units

where the speed of light is one. The entropy $\delta S_{\rm rad}$ of this the shell of radiation is given by the thermodynamic formula

$$\begin{aligned}\delta S_{\rm rad} &= N\int_0^\infty \frac{\mathrm{d}\omega}{2\pi}\left[(1+\langle N_\omega\rangle)\log(1+\langle N_\omega\rangle)-\langle N_\omega\rangle\log\langle N_\omega\rangle\right]\delta u \\ &= \frac{\pi N T}{6}\delta u,\end{aligned} \tag{3.2.7}$$

where N is the number of fields which make up the radiation. The thermal entropy of all the radiation emitted since the formation of the black hole at $u=0$ and a later time u is therefore

$$S_{\rm rad} = \frac{\pi N}{6}\int_0^u T(u')\mathrm{d}u'. \tag{3.2.8}$$

This also follows from an exact CFT calculation. The entropy of an interval of radiation collected in an interval with coordinates (U_1, v_1) and (U_2, v_2), is

$$S_{\rm QFT} = \frac{N}{6}\log\frac{-(U_1-U_2)(v_1-v_2)}{\Omega_1\Omega_2}, \tag{3.2.9}$$

where

$$\Omega = \sqrt{-2\pi T(u)U}, \tag{3.2.10}$$

is the contribution from the Weyl associated to the endpoints which arises from rescaling the Minkowski metric to the metric $-\mathrm{d}U\mathrm{d}v$. This formula is valid for any CFT but for simplicity we can take it to be a theory of N massless Dirac fermions. Ignoring the ultraviolet cutoff and the subleading contribution from the ingoing modes, one finds

$$S_{\rm QFT} \approx \frac{N}{6}\log\frac{\sinh\pi\int_{u_1}^{u_2}T(u)\mathrm{d}u}{\pi\sqrt{T(u_1)T(u_2)}} \to \frac{\pi N}{6}\int_{u_1}^{u_2}T(u)\mathrm{d}u. \tag{3.2.11}$$

Up to the terms that we are ignoring, the first equality is only approximate because we have used the approximate solution (3.2.4) for the function F. The limit on the RHS is the thermodynamic limit where interval is large compared with the thermal scale.

3.2.4 Entanglement across the horizon

The state of the outgoing modes in the Unruh state $|U\rangle$, expressed in terms of the vacuum of the outgoing modes in the vacuum state $|0\rangle$ associated to an asymptotic observer, is a two-mode squeezed state,

$$|U\rangle = \bigotimes_\omega \frac{1}{\sqrt{Z_\omega}}\exp\left[e^{-\omega/2T}a(f_\omega)^\dagger\otimes a(\tilde{f}^*_\omega)^\dagger\right]|0\rangle\,, \tag{3.2.12}$$

where Z_ω is a normalisation. In this expression f_ω is a mode with positive frequency, with respect to the frame of an asymptotic observer, with support outside the black

hole, that is, for $U < 0$, and $\tilde{f}_\omega(U) = f_\omega(-U)$ is partner mode with support inside the black hole, that is, for $U > 0$. These two modes are thermally entangled in Unurh state.

Since the modes f_ω have definite frequency, they are completely delocalised in the outgoing null coordinate u. However, we can construct localised wavepackets by constructed by taking superpositions of positive frequency modes. Since the characteristic frequency is $\mathcal{O}(1/T)$, we should be able to define a basis of wavepackets which are localised over the scale $\Delta u = \mathcal{O}(1/T)$. This intuition tells us that the Hawking radiation collected in a null interval $R = [u_1, u_2]$,which is large compared to the thermal scale, will be thermally entangled with a set modes collected in a null interval behind the horizon and related to R by reflection across the horizon $U \to -U$.

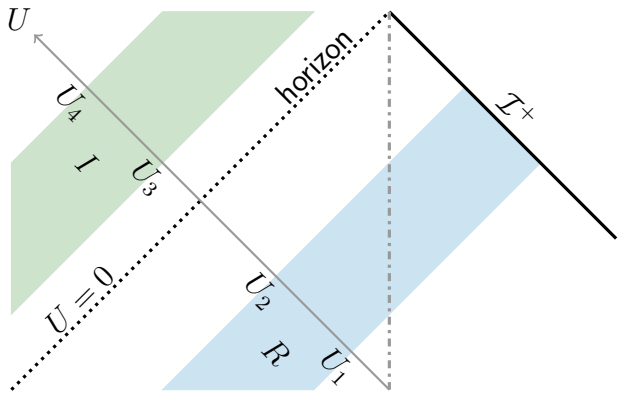


Figure 3.2: The interval $R = [U_1, U_2]$ collects a subset of the outgoing Hawking radiation whilst the interval $I = [U_3, U_4]$ behind the horizon collects a subset of the modes which are entangled with the Hawking radiation.

To make this intuition more precise, consider the intervals $R = [U_1, U_2]$, which collects a subset of the outgoing Hawking radiation, and $I = [U_3, U_4]$, which collects a subset of the modes behind the horizon which are entangled with the Hawking radiation, as show in figure 3.2. Suppose that the interval I lies near the horizon, in the sense that both U_3V_3 and U_4V_4 are much less than one. Now suppose that

$$|U_1| \gg U_4 \gg |U_2| \gg U_3. \tag{3.2.13}$$

In this limit the entropies of each of the individual intervals I and R will be well approximated by the thermodynamic formula, as seen in the previous section. We then ask: what is their combined entropy? That is, what is $S_{\text{QFT}}(I \cup R)$?

For simplicity, we will restrict to the case that the matter sector consists of N massless Dirac fermions, as this is the only CFT for which the multi-interval entropy is known analytically [79]. Ignoring the ultraviolet cutoff and the contribution from the

ingoing modes, we find[5]

$$S_{\text{QFT}}(I \cup R) = \frac{N}{6} \log \frac{U_2 - U_1}{\sqrt{U_1 U_2}} + \frac{N}{6} \log \frac{U_4 - U_3}{\sqrt{U_3 U_4}} + \frac{c}{6} \log \frac{(U_4 - U_1)(U_3 - U_2)}{(U_4 - U_2)(U_3 - U_1)}. \quad (3.2.14)$$

Using (3.2.13), we find

$$\begin{aligned} S_{\text{QFT}}(I \cup R) &\approx \frac{N}{12} \log \frac{-U_2}{U_3} + \frac{N}{12} \log \frac{-U_1}{U_4} \\ &= S_{\text{rad}}([u_1, \tilde{u}_4]) + S_{\text{rad}}([u_2, \tilde{u}_3]). \end{aligned} \quad (3.2.15)$$

Here we used the coordinate $\tilde{u}$ to label the points behind the horizon, which is defined by

$$U = \exp\left(-2\pi \int_0^{\tilde{u}} \mathrm{d}\tilde{u}'\, T(\tilde{u}')\right). \quad (3.2.16)$$

We can also think of $\tilde{u}$ as the outgoing null coordinate associated to the endpoints of I after being reflected across the horizon under $U \to U$ or $\tilde{u} \to u = \tilde{u}$. Notice that we also get (3.2.15) for a general CFT since the relative magnitudes (3.2.13) is similar to being in the OPE limit $U_1 \to U_4$, $U_2 \to U_3$, in the sense that the cross ratio

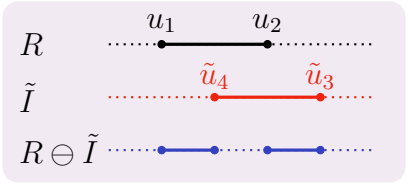
$$\frac{(U_2 - U_1)(U_4 - U_3)}{(U_4 - U_2)(U_3 - U_1)} \to 1, \quad (3.2.17)$$

and the form of the entropy is fixed universally by conformal symmetry. Related ideas will play an important role in chapter 5.

The intuition behind (3.2.15) is that the subset of modes in $[\tilde{u}_4, u_2]$ in R are entangled with modes in I so do not contribute to the entropy of $I \cup R$. On the other hand, the subset of modes in $[u_2, \tilde{u}_3]$ add to the entropy. The result (3.2.15) can be neatly packaged in terms of the symmetric difference $I \ominus R = I \cup R - I \cap R$,

$$S_{\text{QFT}}(I \cup R) \approx S_{\text{rad}}(\tilde{I} \ominus R). \quad (3.2.18)$$

Here $\tilde{I}$ is the image of I under the reflection $U \to -U$ across the horizon. We can visualise the overlapping sets of modes using the diagram:



[5]There is a subtlety that we are sweeping under the rug here. The Weyl anomaly associated to the endpoints of R arise from the rescaling $-\mathrm{d}u\mathrm{d}v \to -\mathrm{d}U\mathrm{d}v$ so contribute as $\sqrt{-U}$. On the other hand, the Weyl anomaly associated to the endpoints of I, which are assumed to lie near the horizon, arise from the rescaling $-4\mathrm{d}U\mathrm{d}V \to -\mathrm{d}U\mathrm{d}v$ so contribute as $1/\sqrt{V}$, not $\sqrt{U}$. Serendipitously, however, it turns out that the endpoints of I, when determined by extremising the generalised entropy, will have Kruskal coordinates related by $U \sim 1/V$.

Note that for $\tilde{I} = R$, $S_{\text{QFT}}(\tilde{R} \cup R) \approx 0$. That is, the state on $\tilde{R} \cup R$ is approximately pure.

3.3 Islands in the stream

In this section we infer the existence of a class of extrema of the generalised entropy using intuitive arguments. The detailed proof is relegated to appendix B.1.

The first assumption is that the putative QESs lie close to the horizon. This can be shown to be self consistent by the following argument. Since the ingoing modes are in the vacuum associated to the frame of an asymptotic observer (that is, an observer who defines positive frequency with respect to boundary time t), at leading order, they contribute to the QFT entropy only through the Weyl factor in the AdS region. Hence, the dependence of the generalised entropy on each of the ingoing null coordinates V_a of the QESs, labelled by the index a, is particularly simple

$$S_{\text{gen}} = (S_{\text{BH}}(v_a) - S_0)(1 - 2U_a V_a) + \frac{N}{12} \log V_a + \cdots \tag{3.3.1}$$

Here the first term is the area term evaluated near the horizon $U = 0$, the second term arises from the Weyl factor associated to the outgoing null coordinates of the QES, and the $\ldots$ denote terms which do not depend on V_a or are subleading. The negative flux of energy across the horizon[6]

$$\frac{\mathrm{d}E}{\mathrm{d}t} = \frac{c}{24\pi}\{F, t\} \approx -\frac{\pi N}{12} T(t)^2, \tag{3.3.2}$$

together with the first law imply

$$\mathrm{d}S_{\text{BH}} = \frac{1}{T}\mathrm{d}E = -\frac{N}{24V}\mathrm{d}V. \tag{3.3.3}$$

Extremising the generalised with respect to V_a then gives

$$U_a V_a = \frac{N}{48} \frac{1}{S_{\text{BH}}(v_a) - S_0}, \tag{3.3.4}$$

and so when k is much smaller than the thermal scale, $U_a V_a \ll 1$ and so the QES does indeed lie close to the horizon.

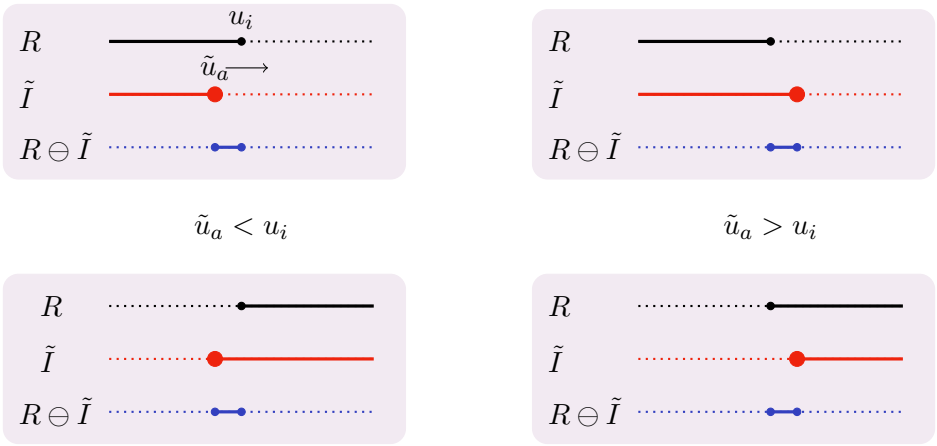
A subtle point regarding (3.3.4) is that the leading behaviour of U_a and V_a cancel in the combination $U_a V_a$, so (3.3.4) is actually a subleading effect. We will see that

[6] In the Unruh state $\langle T_{UU} \rangle = 0$ and $\langle :T_{vv}: \rangle = 0$, as discussed in section 2.2. Hence, the only contribution to the energy balance equation (2.3.7) comes from the negative flux of energy across the horizon $\langle T_{vv} \rangle$, which is determined by the conformal anomaly.

this subleading term is responsible for the scrambling time, while at leading order we have $\tilde{u}_a = v_a$. This means that the area contribution to the generalised entropy can be written as the Bekenstein-Hawking entropy evaluated at the ingoing time $v_a = \tilde{u}_a$. Notice that $\tilde{u}_a$ is the outgoing null coordinate of the "mirror" $\partial\tilde{I}$, which is related to the QES ∂I by a reflection across the horizon. This gives

$$S_{\text{gen}}(I \cup R) = \sum_{\substack{\text{disconnected} \\ \text{components} \\ \partial\tilde{I}_k \text{ of } \partial\tilde{I}}} S_{\text{BH}}(u_{\partial\tilde{I}_k}) + S_{\text{rad}}(\tilde{I} \ominus R). \tag{3.3.5}$$

The remaining task is to extremise the generalised entropy with respect to the outgoing null coordinates U_a of the QESs. Consider the variation of the generalised entropy as one of the endpoints $\tilde{I}$ varies in a neighbourhood of an endpoint of ∂R. A key assumption being made here is the subsets of radiation and spaces between them are large compared with the thermal scale so that the thermodynamic formula for the entropy (3.1.6) holds. There are four possible scenarios, the first two with $\tilde{u}_a$ increasing through u_i, where the index i labels the endpoints of ∂R, from left to right, are:

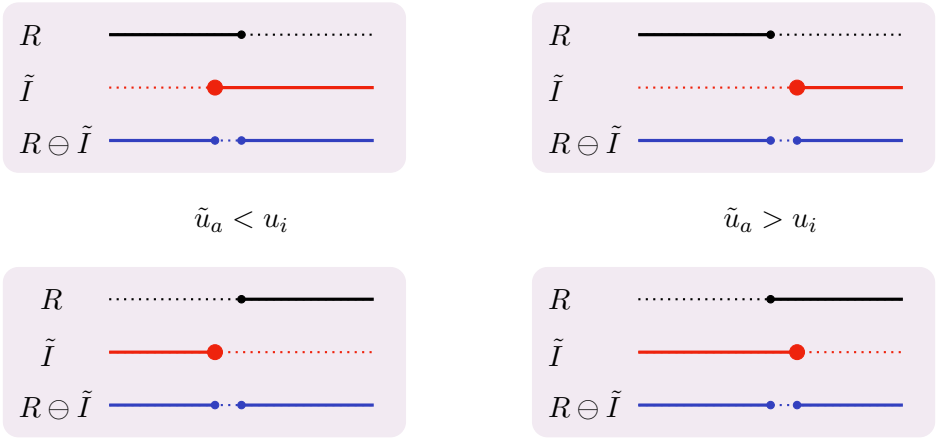


The QES contributes $\mathcal{S}_a := S_{\text{BH}}(\tilde{u}_a)$ to the entropy whilst the contribution to S_{QFT} comes from the blue region as is given by $S_{\text{rad}}(I \ominus R) = 2|\mathcal{S}_a - \mathcal{S}_i|$. Hence as $\tilde{u}_a$ increases

$$\frac{\partial S_{\text{gen}}}{\partial \tilde{u}_a} = \begin{cases} 3\mathcal{S}'_a & \text{for } \tilde{u}_a < u_i, \\ -\mathcal{S}'_a & \text{for } \tilde{u}_a > u_i, \end{cases} \tag{3.3.6}$$

for either of the two scenarios. Here $'$ denotes a derivative with respect to $\tilde{u}_a$. Since the area of the evaporating black hole decreases $\mathcal{S}' < 0$, the generalised entropy has a minimum at $\tilde{u}_a = u_i$.

The other two possible scenarios are:



For these cases, as $\tilde{u}_a$ increases, we have

$$\frac{\partial S_{\text{gen}}}{\partial \tilde{u}_a} = \begin{cases} -\mathcal{S}'_a & \text{for } \tilde{u}_a < u_i, \\ 3\mathcal{S}'_a & \text{for } \tilde{u}_a > u_i. \end{cases} \tag{3.3.7}$$

So for these cases the generalised entropy has a maximum at $\tilde{u}_a = u_i$. In the detailed analysis, presented in appendix B.1, we find that the generalised entropy behaves smoothly as $\tilde{u}_a$ moves across u_i, in contrast to the discontinuous behaviour seen here.

There exists a class of extrema where each element of $\partial\tilde{I}$ is mapped to a unique element of ∂R with $\tilde{u}_a \approx u_i$. As we have seen, only a subset of the extrema will actually be minima of the generalised entropy and so have a chance at contributing to $S(R)$. In appendix B.1 we prove this in detail and also compute the leading order corrections to the coordinates of the QESs:

$$\begin{aligned} u_a &= u_i - \frac{1}{2\pi T(u_i)} \log \lambda_a, \\ v_a &= u_i - \frac{1}{2\pi T(u_i)} \log \frac{48\lambda_a (S_{\text{BH}}(u_i) - S_0)}{N}, \end{aligned} \tag{3.3.8}$$

with λ_a equal to $\frac{1}{3}$ or 3 for the for minimum or maximum cases above, respectively. The difference

$$t_{\text{scr}} = u_i - v_a = \frac{1}{2\pi T(u_i)} \log \frac{48\lambda_a (S_{\text{BH}}(u_i) - S_0)}{N}, \tag{3.3.9}$$

is identified with the scrambling time, which is a subleading effect in the adiabatic limit.

3.4 Entropy inequalities

In this section we describe some applications of the islands in the stream formula (3.1.7) and show that constraints from unitarity and various entropy inequalities, including the

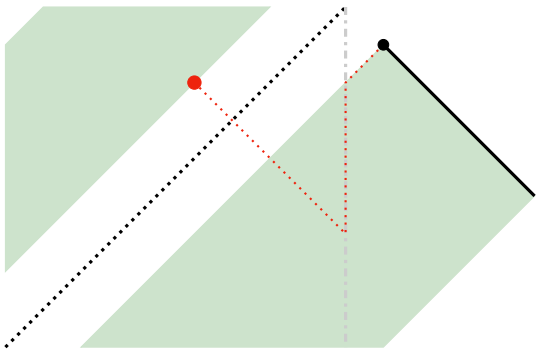


Figure 3.3: The relationship between a point in ∂R, shown in black, and a QES ∂I, shown in red. At leading order in the adiabatic limit, the outgoing null coordinate U of the QES is determined by the reflection across the horizon $U \to -U$ of the outgoing null coordinate of the point in ∂R. The ingoing null coordinate v of the QES is equal to the $u - t_{\text{scr}}$ where u is the outgoing null coordinate of the point in ∂R.

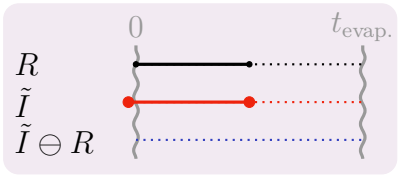
Araki-Lieb inequality, subadditivity, and strong subadditivity, are obeyed. For simplicity, in this section we will assume that the extremal entropy S_0 is negligible and ignore its contribution. This restriction is not necessary and its effect can straightforwardly be reintroduced.

3.4.1 Page curve

To derive the Page curve, we consider an interval $R = [0, u]$ which captures all the Hawking radiation emitted by the black hole from the time of formation $u = 0$ up to time u. There are two possible islands. The first is the empty island $I = \emptyset$, the analogue of Hawking's result, which has generalised entropy[7]

$$S_{\text{gen}}(R) = S_{\text{rad}}(R) = 2(\mathcal{S}_0 - \mathcal{S}_u). \tag{3.4.1}$$

The second island is $I = \tilde{R}$, or equivalently $\tilde{I} = R$, with one of the QESs located at the point just before the black hole was formed by an infalling shockwave:



Since $\tilde{I} \ominus R = \emptyset$, the generalised entropy with this island is

$$S_{\text{gen}}(I \cup R) = S_{\text{BH}}(u) = \mathcal{S}_u. \tag{3.4.2}$$

[7]The parameter $\mathcal{S}_0 = S_{\text{BH}}(0)$ should not be confused with the extremal entropy S_0.

The entropy of R is then given by

$$S(R) = \min\{2(\mathcal{S}_0 - \mathcal{S}_u), \mathcal{S}_u\}. \tag{3.4.3}$$

For early times the empty island dominates but at the Page time

$$\mathcal{S}_{u_{\text{Page}}} = \frac{2}{3}\mathcal{S}_0, \tag{3.4.4}$$

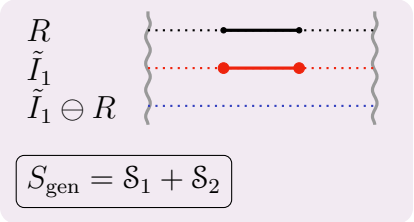
there is a transition from the empty island to the nonempty island $I = \tilde{R}$.

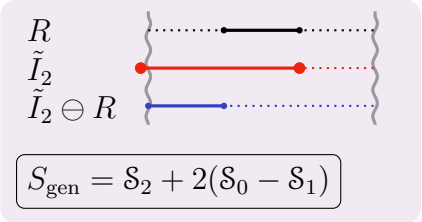
3.4.2 Single interval

Consider a single interval $R = [u_1, u_2]$. Again, there is an empty island with generalised entropy

$$S_{\text{gen}}(R) = 2(\mathcal{S}_{u_1} - \mathcal{S}_{u_2}). \tag{3.4.5}$$

For this case there are two possible nonempty islands:





There is a competition between the three islands

$$S(R) = \min\{2(\mathcal{S}_1 - \mathcal{S}_2), \mathcal{S}_1 + \mathcal{S}_2, \mathcal{S}_2 + 2(\mathcal{S}_0 - \mathcal{S}_1)\}. \tag{3.4.6}$$

The final island can only contribute if u_1 is less than the Page time, that is,

$$\mathcal{S}_1 > \frac{2}{3}\mathcal{S}_0. \tag{3.4.7}$$

If u_1 is greater than the Page time then the empty island contributes only when the interval $R = [u_1, u_2]$ is sufficiently small, specifically

$$\Delta_R \coloneqq 3\mathcal{S}_2 - \mathcal{S}_1 > 0. \tag{3.4.8}$$

3.4.3 Unitarity

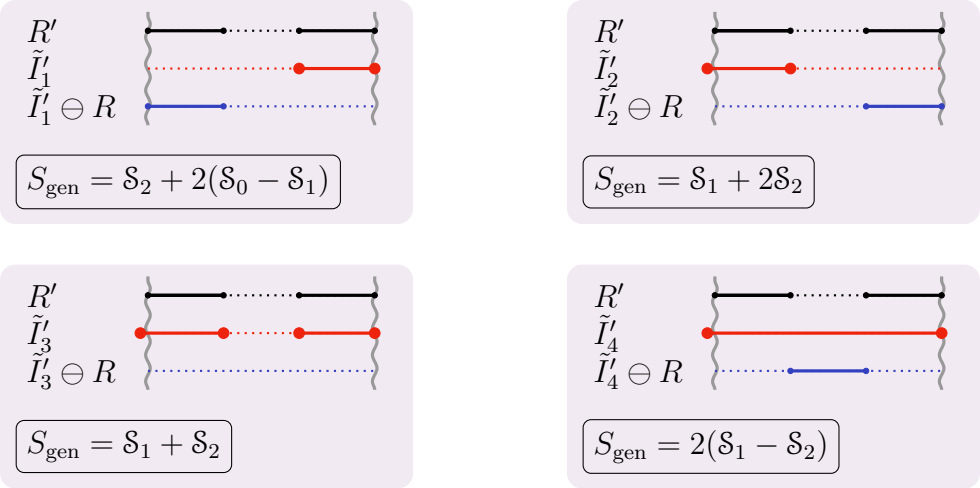
Consider the single interval $R = [u_1, u_2]$ studied in the previous section and let R' denote the complementary region:

$$R' = [0, u_1] \cup [u_2, t_{\text{evap}}]. \tag{3.4.9}$$

Here t_{evap} denotes the timescale for evaporation, which for the black hole in JT gravity is equal to ∞. There are various islands which can contribute to the entropy of the complementary region R'. Firstly, there is the empty island with generalised entropy

$$S_{\text{gen}}(R') = 2(\mathcal{S}_0 - \mathcal{S}_1 + \mathcal{S}_2). \tag{3.4.10}$$

There are four possible islands:



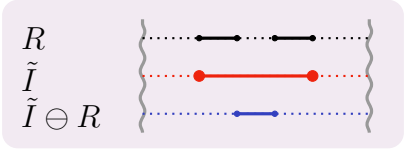
The empty island can never contribute since $S_{\text{gen}}(R') > S_{\text{gen}}(I'_1 \cup R')$. In addition, the island I'_2 can never contribute since $S_{\text{gen}}(I'_2 \cup R') > S_{\text{gen}}(I'_3 \cup R')$. If the total state of the system is pure, which must be the case if the system evolved unitarily from an initial pure state, we must have $S(R) = S(R')$. Indeed, we find that this is the case since there is a perfect matching of the islands that can contribute:

$$S_{\text{gen}}(R) = S_{\text{gen}}(I'_4 \cup R'), \qquad S_{\text{gen}}(I_1 \cup R) = S_{\text{gen}}(I'_3 \cup R'), \qquad S_{\text{gen}}(I_2 \cup R) = S_{\text{gen}}(I'_1 \cup R'). \tag{3.4.11}$$

This provides a nontrivial test of unitarity in the sense described above. It is interesting to note that for all these cases the reflections of the islands are complementary, e.g. $\tilde{I}_1 \cap \tilde{I}'_3 = \emptyset$ whilst $\tilde{I}_1 \cup \tilde{I}'_3 = [0, t_{\text{evap}}]$.

3.4.4 Mutual information of two intervals

Consider two disjoint intervals R_1 and R_2 and let $R = R_1 \cup R_2$. The mutual information $I(R_1 : R_2) = S(R_1) + S(R_2) - S(R_1 \cup R_2)$ will only be nonvanishing when there is an island that stretches between R_1 and R_2:



In order to compute the mutual information we need to know whether there is an island contributing to R_1 or R_2. For simplicity, let us assume that u_1 is greater than the Page time, that is, the inequality (3.4.7) is not satisfied. The final result remains unchanged if we relax this condition. The condition for the empty island to contribute to the entropy of R_1 and R_2 is determined by the conditions $\Delta_{R_1} > 0$ and $\Delta_{R_2} > 0$, respectively, where Δ was defined in (3.4.8). Let P denote the interval between R_1 and R_2, then there are four possible cases:

1. $\tilde{I}_1 = \emptyset$ and $\tilde{I}_2 = \emptyset$: $I(R_1 : R_2) = \max(0, \Delta_P - \Delta_{R_1} - \Delta_{R_2})$.
2. $\tilde{I}_1 = \emptyset$ and $\tilde{I}_2 = R_2$: $I(R_1 : R_2) = \max(0, \Delta_P - \Delta_{R_1})$.
3. $\tilde{I}_1 = R_1$ and $\tilde{I}_2 = \emptyset$: $I(R_1 : R_2) = \max(0, \Delta_P - \Delta_{R_2})$.
4. $\tilde{I}_1 = R_1$ and $\tilde{I}_2 = R_2$: $I(R_1 : R_2) = \max(0, \Delta_P)$.

Remarkably we can combine all these separate cases into a single expression which automatically incorporates the effect of which island contributes to R_1 and R_2 on their own:

$$I(R_1 : R_2) = \max(0, \Delta_P - \max(\Delta_{R_1} + \Delta_{R_2}, \Delta_{R_1}, \Delta_{R_2}, 0)). \tag{3.4.12}$$

Notice that that the mutual information is manifestly positive. This property is also known as subadditivity of entropy.

3.4.5 Araki-Lieb inequality

The Araki-Lieb inequality $S(R_1 \cup R_2) \geq |S(R_1) - S(R_2)|$ [83] can be rephrased as an upper bound on the mutual information

$$I(R_1 : R_2) \leq 2 \min(S(R_1), S(R_2)). \tag{3.4.13}$$

We will prove this bound using the setup of the previous section.[8] From (3.4.12) we see that to compute the mutual information we need to take minus the maximum among several possibilities. Assuming the mutual information is non vanishing, $I(R_1 : R_2) \neq 0$, it is enough to prove that $2S(R_{1,2})$ is larger than one of these quantities to prove that the Araki-Lieb inequality is satisfied.

First consider the subsystem R_1. Since the entropy of an interval is always larger than half of the generalised entropy with an empty island, $S(R_1) \geq \frac{1}{2} S_{\text{gen}}(R_1)$, we have the following chain of inequalities:

$$2S(R_1) \geq S_{\text{gen}}(R_1) \geq \Delta_P - \Delta_{R_1} \geq I(R_1 : R_2). \tag{3.4.14}$$

For R_2 it is more convenient to consider the various generalised entropies separately:

$$2S_{\text{gen}}(R_2) \geq \Delta_P - \Delta_{R_2}, \qquad 2S_{\text{gen}}(I_2 \cup R_2) \geq \Delta_P \qquad \Rightarrow \qquad 2S(R_2) \geq I(R_1, R_2). \tag{3.4.15}$$

This proves that the islands in the stream formula (3.1.7) is consistent with the Araki-Lieb inequality.

3.4.6 Monogamy of mutual information

Given a tripartite system with subsystems R_1, R_2, and R_3, the conditional mutual information

$$I(R_1 : R_2|R_3) = S(R_1 \cup R_3) + S(R_2 \cup R_3) - S(R_3) - S(R_1 \cup R_2 \cup R_3), \tag{3.4.16}$$

is meant to quantify the correlations between R_1 and R_2 from the point of view of R_3. Strong subadditivity of entropy may be rephrases as the nonnegativity of conditional mutual information. We will prove a more stringent inequality, known as monogamy of mutual information (MMI):[9]

$$I(R_1 : R_2|R_3) \geq I(R_1 : R_2). \tag{3.4.17}$$

Whilst strong subadditivity holds for any quantum system, MMI need not hold. If it does hold, it suggests that the correlations between R_1 and R_2 are mostly quantum. This follows from two facts. Firstly, minimising $I(R_1 : R_2|R_3)$ over all possible R_3 provides a measure of entanglement between R_1 and R_2 [85]. On the other hand, the mutual information $I(R_1 : R_2)$ provides an upper bound on the correlations, both classical and quantum, between R_1 and R_2. Hence, MMI is telling us that there is no room for classical correlations.

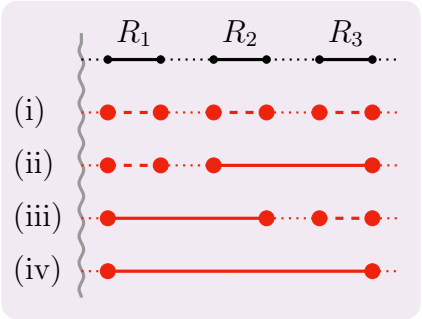
[8] In particular, we assume u_1 is greater than the Page time. Again, this condition can be relaxed at the expense of complicating the analysis although without changing the conclusion.

[9] A more general argument of strong subadditivity is given in [84]. The authors also investigate MMI and argue that it is obeyed if the bulk QFT obeys MMI. This is the case for the entropy of thermal radiation.

In order to prove (3.4.17), it is convenient to rewrite the inequality in a way that is totally symmetric in R_1, R_2, and R_3:

$$S(R_1)+S(R_2)+S(R_3)+S(R_1\cup R_2\cup R_3) \leq S(R_1\cup R_2)+S(R_1\cup R_3)+S(R_2\cup R_3). \quad (3.4.18)$$

We assume that the first subset R_1 is subject to the constraint $\mathcal{S}_1 < \frac{2}{3}\mathcal{S}_0$. As usual, this condition can be relaxed at the expense of complicating the analysis although without changing the conclusion. We now divide the proof into cases depending on how $S(R_1 \cup R_2 \cup R_3)$ factorises as governed by which island contributes to the entropy. There are four possible cases:



where the dashed segments underneath a subsystem R_j indicate an empty island or one with a component $\tilde{R}_j$.

(i) In this case, $S^{(\mathrm{i})}_{\mathrm{gen}} = S(R_1) + S(R_2) + S(R_3)$ and $S(R_i \cup R_j) = S(R_i) + S(R_j)$. Hence (3.4.18) is trivially satisfied as an equality.

(ii) In this case $S^{(\mathrm{ii})}_{\mathrm{gen}} = S(R_1) + S(R_2 \cup R_3)$ and it follows that $S(R_1 \cup R_2) = S(R_1) + S(R_2)$ and $S(R_1 \cup R_3) = S(R_1) + S(R_3)$. Hence (3.4.18) is satisfied as an equality.

(iii) The proof for this case is analogous to the proof for case (ii).

(iv) This is the trickiest case since $S^{(\mathrm{iv})}_{\mathrm{gen}}$ does not factorise and we have to consider the various subcases where an island contributes to $S(R_i \cup R_j)$. Let $\mathcal{U}$ the set of pairs that receive a contribution from an island. Of the eight sets of pairs $\mathcal{U} = \{R_1 \cup R_2, R_2 \cup R_3\}$ cannot actually occur as this would imply $S(R_1 \cup R_3)$ also receives a contribution from an island. Pairs $\{R_i, R_j\}$ that are not in $\mathcal{U}$ satisfy $S(R_i \cup R_j) = S(R_i) + S(R_j)$. Consider now the remaining seven subcases:

(iv.1) For $\mathcal{U} = \emptyset$, (3.4.18) reads

$$S^{(\mathrm{iv})}_{\mathrm{gen}} \overset{?}{\leq} S(R_1) + S(R_2) + S(R_3) = S^{(\mathrm{i})}_{\mathrm{gen}}, \quad (3.4.19)$$

which is satisfied since we are assuming $S^{(\mathrm{iv})}_{\mathrm{gen}}$ has minimum generalised entropy.

(iv.2) For $\mathcal{U} = \{R_1 \cup R_2\}$, (3.4.18) reads

$$S_{\text{gen}}^{(\text{iv})} \overset{?}{\leq} S(R_1 \cup R_2) + S(R_3) = S_{\text{gen}}^{(\text{iii})}, \tag{3.4.20}$$

which is satisfied since we are assuming $S_{\text{gen}}^{(\text{iv})}$ has minimum generalised entropy.

(iv.3) The proof for the case that $\mathcal{U} = \{R_2 \cup R_3\}$ is analogous to the proof for case (ii.2).

(iv.4) For $\mathcal{U} = \{R_1 \cup R_3\}$, (3.4.18) reads

$$S_{(\text{iv})} \overset{?}{\leq} S(R_2) + S(R_1 \cup R_3). \tag{3.4.21}$$

When $S(R_1 \cup R_3)$ receives a contribution from an island, a useful identity is

$$S(R_1 \cup R_3) = S_{\text{gen}}^{(\text{iv})} + S_{\text{gen}}(R_2). \tag{3.4.22}$$

Using this relation (3.4.21) becomes

$$0 \overset{?}{\leq} S(R_2) + S_{\text{gen}}(R_2), \tag{3.4.23}$$

which is clearly satisfied since the RHS is a sum of nonnegative quantities.

(iv.5) For $\mathcal{U} = \{R_1 \cup R_2, R_1 \cup R_3\}$, using the relation (3.4.22), (3.4.18) reads

$$S(R_1) - S_{\text{gen}}(R_2) \overset{?}{\leq} S(R_1 \cup R_2). \tag{3.4.24}$$

The Araki-Lieb inequality and the fact that $S(R_2) \leq S_{\text{gen}}(R_2)$ then lead to

$$S(R_1 \cup R_2) \geq S(R_1) - S(R_2) \geq S(R_1) - S_{\text{gen}}(R_2), \tag{3.4.25}$$

which implies that (3.4.24) holds.

(iv.6) The proof for the case that $\mathcal{U} = \{R_1 \cup R_3, R_2 \cup R_3\}$ is analogous to the proof for case (ii.5).

(iv.7) For $\mathcal{U} = \{R_1 \cup R_3, R_2 \cup R_3, R_3 \cup R_1\}$, using the relation (3.4.22), (3.4.18) reads

$$\begin{aligned} 0 \overset{?}{\leq} &[S(R_1 \cup R_2) - S(R_1) + S(R_2)] \\ &+ [S(R_2 \cup R_3) - S(R_2) + S(R_3)] + [S_{\text{gen}}(R_2) - S(R_2)]\,. \end{aligned} \tag{3.4.26}$$

Each of the three terms in the parentheses are nonnegative: the first two as a consequence of the Araki-Lieb inequality, and the last by definition of a QES.

3.5 Adding a greybody factor

In this section we consider the possibility of a nontrivial greybody factor. Greybody factors are a generic feature of black holes in higher dimensions which arise from the fact that Hawking radiation must tunnel through an effective potential barrier around the black hole. This means that some of the modes will be reflected back into the black hole and so there is a mixing of ingoing and outgoing modes. In particular, only some of the radiation escapes out to infinity and the spectrum of Hawking radiation gets modified from a blackbody to a greybody,

$$\langle N_\omega^{\mathbb{T}} \rangle = \frac{\Gamma(\omega)}{e^{\omega/T} - 1}. \tag{3.5.1}$$

In JT gravity we can mimic the effect of the potential barrier by picking certain boundary conditions for the matter quantum field theory along the AdS-Minkowski interface so that modes are only partly transmitted. We can think of this as a toy model describing the two-dimensional effective theory of black hole in higher dimensions but where the near-horizon region of the black hole, the zone, which corresponds to the nearly AdS_2 factor, has been glued to the asymptotic region, which corresponds to the half Minkowski space. Since this is a toy model we can basically engineer any greybody factor we like. For example, for a free scalar field which satisfies the wave equation, we can pick boundary conditions so that an incoming mode with frequency ω is partially transmitted at the boundary $u = v$:

$$\varphi_\omega(u, v) = \begin{cases} e^{-\mathrm{i}\omega u} + \mathbb{R}(\omega) e^{-\mathrm{i}\omega v} & \text{for } u < v, \\ \mathbb{T}(\omega) e^{-\mathrm{i}\omega u} & \text{for } u > v. \end{cases} \tag{3.5.2}$$

The greybody factor is then defined as the transmission probability

$$\Gamma(\omega) \coloneqq |\mathbb{T}(\omega)|^2 = 1 - |\mathbb{R}(\omega)|^2. \tag{3.5.3}$$

Since the transmitted radiation carries energy away from the black hole, its energy decreases as

$$\frac{\mathrm{d}E}{\mathrm{d}t} = -N \int_0^\infty \frac{\mathrm{d}\omega}{2\pi} \frac{\omega \Gamma(\omega)}{e^{\omega/T} - 1}. \tag{3.5.4}$$

The dependence of the temperature T on time then follows from the relation between the energy E of the black hole and its temperature T. The first law also implies that the Bekenstein-Hawking entropy decreases since

$$\frac{\mathrm{d}S_{\mathrm{BH}}}{\mathrm{d}t} = \frac{1}{T} \frac{\mathrm{d}E}{\mathrm{d}t}. \tag{3.5.5}$$

In the presence of a nontrivial greybody factor the radiation which escapes out to infinity is no longer thermal. However, in the zone, the outgoing modes are still in the Unruh state

$$|U\rangle = \bigotimes_\omega \frac{1}{\sqrt{Z_\omega}} \sum_{n=0}^{1 \text{ or } \infty} e^{-n\omega/2T} \, |n\rangle_{\mathrm{int}}^\omega \otimes |n\rangle_{\mathrm{ext}}^\omega \, . \tag{3.5.6}$$

As before, Z_ω a normalisation, and $|n\rangle^\omega_{\text{int}}$ and $|n\rangle^\omega_{\text{ext}}$ denote a state of definite frequency ω in the interior and exterior, respectively. The sum goes up to 1 for a fermion and ∞ for a scalar. In the exterior, the modes can be split into a sum of transmitted $\mathbb{T}$ and reflected $\mathbb{R}$ modes. This gives

$$|U\rangle = \bigotimes_\omega \frac{1}{\sqrt{Z_\omega}} \sum_{n=0}^{1 \text{ or } \infty} e^{-n\omega/2T} \sum_{k=0}^{n} \sqrt{\binom{n}{k} \Gamma(\omega)^{n-k}(1-\Gamma(\omega))^k}\, |n\rangle^\omega_{\text{int}} \otimes |n-k\rangle^\omega_{\mathbb{T}} \otimes |k\rangle^\omega_{\mathbb{R}}\,. \tag{3.5.7}$$

The reduced state for the transmitted modes is obtained by tracing out both the reflected modes and the interior modes

$$\rho_{\mathbb{T}} = \bigotimes_\omega \frac{1}{Z_\omega} \sum_{n=0}^{1 \text{ or } \infty} e^{-n\omega/T} \sum_{k=0}^{n} \binom{n}{k} \Gamma(\omega)^{n-k}(1-\Gamma(\omega))^k\, |n-k\rangle^\omega_{\mathbb{T}} \langle n-k|^\omega_{\mathbb{T}}\,. \tag{3.5.8}$$

So the eigenvalues $p^{\mathbb{T}}_{n,\omega}$ of the density matrix for a single transmitted mode with frequency ω is

$$\begin{aligned} p^{\mathbb{T}}_{n,\omega} &= \frac{1}{Z_\omega} \left(\Gamma(\omega)e^{-\omega/T}\right)^n \sum_{k=0}^{\infty} \binom{n+k}{n} \left[(1-\Gamma(\omega))e^{-\omega/T}\right]^k \\ &= \frac{1}{Z_\omega} \frac{\left[\Gamma(\omega)e^{-\omega/T}\right]^n}{\left[1-(1-\Gamma(\omega))e^{-\omega/T}\right]^{n+1}}. \end{aligned} \tag{3.5.9}$$

From this, we can deduce the expectation value of the occupation number (3.5.1). We can also express the entropy flux $\mathrm{d}S^{\mathbb{T}}_{\text{rad}}/\mathrm{d}u$ of the transmitted radiation in terms of the expectation value of the occupation number

$$\frac{\mathrm{d}S^{\mathbb{T}}_{\text{rad}}}{\mathrm{d}u} = N \int_0^\infty \frac{\mathrm{d}\omega}{2\pi} \left[(1+\langle N^{\mathbb{T}}_\omega\rangle)\log(1+\langle N^{\mathbb{T}}_\omega\rangle) - \langle N^{\mathbb{T}}_\omega\rangle \log\langle N^{\mathbb{T}}_\omega\rangle\right]. \tag{3.5.10}$$

There is a similar formula for the entropy flux $\mathrm{d}S^{\mathbb{R}}_{\text{rad}}/\mathrm{d}u$ of the reflected radiation.

A simplification occurs in the case that the greybody factor $\Gamma(\omega)$ only depends on the frequency ω and the temperature T. This is the case for a Schwarzschild black hole where there is only a single scale in the problem, the Schwarzschild radius r_s, or equivalently the temperature $T = 1/4\pi r_s$. Such greybody factors can can only depend on ω and T through the dimensionless combination ω/T, that is,

$$\Gamma(\omega) = \tilde{\Gamma}(\omega/T). \tag{3.5.11}$$

This follows because the greybody factor is dimensionless since it's a probability. In this case we can write (3.5.4) as

$$\frac{\mathrm{d}E}{\mathrm{d}t} = -\eta \frac{\pi N T^2}{12}, \tag{3.5.12}$$

where η is the dimensionless parameter

$$\eta = \frac{6}{\pi^2}\int_0^\infty \mathrm{d}x \frac{x\tilde{\Gamma}(x)}{e^x - 1}. \tag{3.5.13}$$

In a similar way, the entropy flux of the transmitted modes (3.5.10) can be expressed as

$$\frac{\mathrm{d}S^{\mathbb{T}}_{\mathrm{rad}}}{\mathrm{d}u} = \alpha_{\mathbb{T}}\frac{\pi N T}{6}, \tag{3.5.14}$$

where $\alpha_{\mathbb{T}}$ is the dimensionless parameter

$$\alpha_{\mathbb{T}} = \frac{3}{\pi^2}\int_0^\infty \mathrm{d}x \left[\left(1 + \frac{\tilde{\Gamma}(x)}{e^x - 1}\right)\log\left(1 + \frac{\tilde{\Gamma}(x)}{e^x - 1}\right) - \frac{\tilde{\Gamma}(x)}{e^x - 1}\log\frac{\tilde{\Gamma}(x)}{e^x - 1}\right]. \tag{3.5.15}$$

There is a similar formula for entropy flux of the reflected modes in terms of a parameter $\alpha_{\mathbb{R}}$ which is given by replacing $\tilde{\Gamma}$ with $1 - \tilde{\Gamma}$ in $\alpha_{\mathbb{T}}$. Notice that for this choice of greybody factor the effect is merely to rescale the energy and entropy fluxes. The greybody coefficient ξ (3.7.18), which measures the ratio of the entropy fluxes of the transmitted radiation and the black hole, is given by the ratio

$$\xi = \frac{2\alpha_{\mathbb{T}}}{\eta}. \tag{3.5.16}$$

For example, for a Schwarzschild black hole and in the s-wave sector:

$$\eta = 0.86, \qquad \alpha_{\mathbb{T}} = 0.82, \qquad \alpha_{\mathbb{R}} = 0.31, \qquad \xi = 1.91. \tag{3.5.17}$$

The upper bound $\xi = 2$ for the greybody coefficient corresponds to transparent boundary conditions, that is, $\Gamma = 1$. In this case $\eta = \alpha_{\mathbb{T}} = 1$ and $\alpha_{\mathbb{R}} = 0$. The lower bound $\xi = 1$ can only be approached as a limit. For example, by taking the greybody factor to only switch on at some high frequency,

$$\Gamma(\omega) = \Theta(\omega/T - x_0), \tag{3.5.18}$$

with $x_0 \gg 1$. In this case

$$\eta \sim \frac{6}{\pi^2}(1 + x_0)e^{-x_0}, \qquad \alpha_{\mathbb{T}} \sim \frac{3}{\pi^2}(2 + x_0)e^{-x_0} \quad \Rightarrow \quad \xi \sim 1 + \frac{1}{x_0}. \tag{3.5.19}$$

So in the limit of large x_0, evaporation becomes very slow and the rate at which the entropy of the black hole is lost is the same as entropy carried away by the transmitted radiation. In this limit, the process of black hole evaporation is thermodynamically reversible, in the sense that the total thermodynamic entropy of the system remains constant, that is,

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(S_{\mathrm{BH}} + S^{\mathbb{T}}_{\mathrm{rad}}\right) \to 0. \tag{3.5.20}$$

3.5.1 Entanglement across the horizon

As we have seen in section 3.2.4, in the thermodynamic limit, the entropy of $I \cup R$, where I is an interval in the interior of the black hole and R is an interval in the exterior which collects the Hawking radiation, is given by

$$S_{\text{QFT}}(I \cup R) \approx S_{\text{rad}}(\tilde{I} \ominus R). \tag{3.5.21}$$

The result is simply stated in terms of the symmetric difference as a consequence of the entanglement structure of modes across the horizon. In this section we will understand how this formula generalises in the presence of a nontrivial greybody factor.

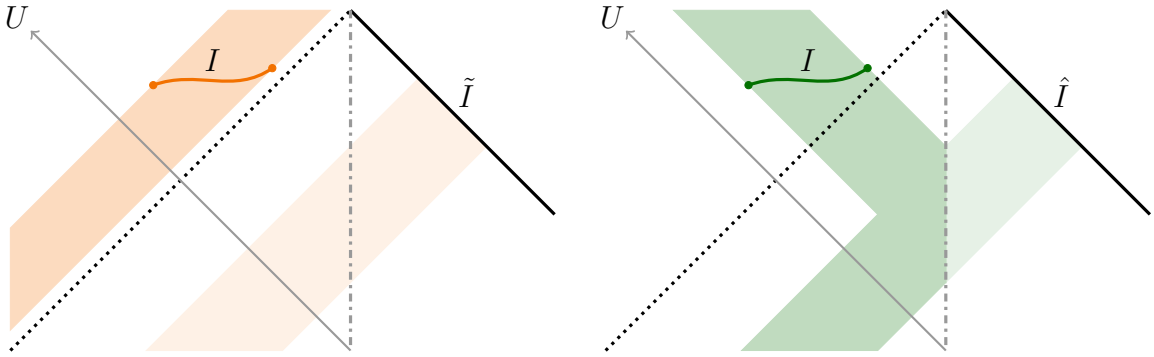


Figure 3.4: On the left: the outgoing modes inside the horizon which pass through I are mapped onto $\mathcal{I}^+$ by a reflection $U \to -U$ followed by a projection. On the right: the ingoing modes inside the horizon which pass through I are mapped onto $\mathcal{I}^+$ by $v \to u$ followed by a projection.

Since the interval I lies behind the horizon, it now also collects modes which are reflected off the potential barrier at the AdS-Minkowski interface. The contribution of these modes to the entropy can be determined by mapping the reflected modes that pass through I to $\mathcal{I}^+$ by $v \to u$. That is, if the range of the ingoing null coordinates v over the interval I is $[v_1, v_2]$, then we define $\hat{I} = [v_1, v_2] \subset \mathcal{I}^+$, as shown in figure 3.4.

In the thermodynamic limit, the entropy $S_{\text{QFT}}(I \cup R)$ depends on the interplay between three sets of modes, as characterised by the subsets $R, \tilde{I}$, and $\hat{I}$. The contributions depend on the tripartite structure of these three sets of modes, as shown in figure 3.5. Explicitly,

$$\begin{aligned} S_{\text{QFT}}(I \cup R) \approx S_{\text{rad}}\left((\tilde{I} \ominus R) \cap (\tilde{I} \ominus \hat{I})\right) + S^{\mathbb{T}}_{\text{rad}}\left((\tilde{I} \ominus R) \cap (\hat{I} \ominus R)\right) \\ + S^{\mathbb{R}}_{\text{rad}}\left((\hat{I} \ominus R) \cap (\hat{I} \ominus \tilde{I})\right), \end{aligned} \tag{3.5.22}$$

with the formulas for $S^{\mathbb{T}}_{\text{rad}}$ and $S^{\mathbb{R}}_{\text{rad}}$ given in the previous section. However, if we work in the limit where the intervals are large compared with the scrambling time, we will

see that in section 3.6 that the endpoints of I, when determined by extremising the generalised entropy, have ingoing and outgoing null coordinates related by

$$\tilde{u} \approx v. \tag{3.5.23}$$

There are important subleading corrections which don't affect the entropy to leading order. The upshot is that to leading order $\tilde{I} = \hat{I}$ and so (3.5.22) simplifies to

$$S_{\text{QFT}}(I \cup R) \approx S^{\mathbb{T}}_{\text{rad}}(\tilde{I} \ominus R). \tag{3.5.24}$$

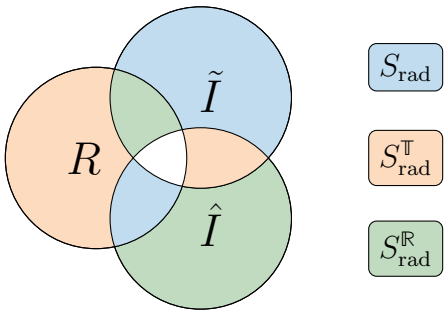


Figure 3.5: An illustration of the tripartite structure of the three sets of modes.

There is a subtlety that we have swept under the rug here. In two-dimensions, the von Neumann entropy has a logarithmic ultraviolet divergence

$$\frac{N}{6} \log \varepsilon, \tag{3.5.25}$$

associated to each of the endpoints of the intervals. The formulas in this section assume the cutoffs ε are measured with respect to the proper distance in the Minkowski metric. This is correct for the endpoints of R, but not for I. The cutoffs associated to the endpoints of I should instead be measured with respect to the AdS_2 metric. Assuming the endpoints of I lie near the horizon, where the metric is $\mathrm{d}s^2 \approx -4\mathrm{d}U\mathrm{d}V$, we need to rescale the cutoff by an amount proportional to

$$\sqrt{\left|\frac{\mathrm{d}U}{\mathrm{d}u}\frac{\mathrm{d}V}{\mathrm{d}v}\right|} \propto \sqrt{|UV|}. \tag{3.5.26}$$

The upshot is that $S_{\text{QFT}}(I \cup R)$ is instead given by

$$S_{\text{QFT}}(I \cup R) \approx S^{\mathbb{T}}_{\text{rad}}(\tilde{I} \ominus R) + S_{\text{Weyl}}(I), \tag{3.5.27}$$

where

$$S_{\text{Weyl}}(I) = \frac{N}{12} \sum_a \log |U_a V_a| \,, \tag{3.5.28}$$

with the sum running over the endpoints of I. The gradient of this term will play an important role when we solve for the QESs in the next section, despite the fact that it turns out to give a subleading contribution to the entropy itself.

3.6 Islands and greybody factors

In this section we extend the proof of the island in the stream formula, given in section 3.3, to the case with a greybody factor. Specifically, we prove the following:

1. The outgoing null coordinate $\tilde{u}_a$ of a point in ∂I is related to the outgoing null coordinate u_i of a point in ∂R by

$$\tilde{u}_a = u_i + \mathcal{O}(1/T). \tag{3.6.1}$$

   The one exception is that there can also be a QES located at the point just before the black hole was formed.

2. The ingoing and outgoing null coordinates, v_a and $\tilde{u}_a$, of a QES are related by

$$v_a = \tilde{u}_a - \varepsilon. \tag{3.6.2}$$

   Here $\varepsilon = \mathcal{O}(T^{-1} \log s^{-1})$ is a subleading correction of the order of the scrambling time and we have defined

$$s \coloneqq \frac{N}{S_{\text{BH}}(u) - S_0} \ll 1. \tag{3.6.3}$$

   Although ε is large since $s \ll 1$ it is still considered subleading as we are working in the limit where the subsets of the radiation and spaces between them are large compared to the scrambling time. Since the temperate of the black hole is slowly varying

$$U_a V_a = \exp\left(-2\pi \int_{\tilde{u}_a - \varepsilon}^{\tilde{u}_a} T(u) \mathrm{d}u\right) \approx e^{-2\pi T(\tilde{u}_a)\varepsilon} \approx s^{\#}, \tag{3.6.4}$$

   for some positive number $\#$. Since $s \ll 1$ this means that the QESs lie close to the horizon.

The generalised entropy $S_{\text{gen}}(I \cup R)$ can then be evaluated at leading order as follows. Since the QESs lie near the horizon, the area term is just the Bekenstein-Hawking entropy $S_{\text{BH}}(v_{\partial I})$ evaluated at each of the QESs. But due to (3.6.2), we can write this as $S_{\text{BH}}(v_{\partial I}) \approx S_{\text{BH}}(u_{\partial \tilde{I}})$. The second term in the generalised entropy can be computed using the arguments in section 3.5.1. But due to (3.6.2), the formula simplifies to (3.5.24), that is $S_{\text{QFT}}(I \cup R) \approx S^{\mathbb{T}}_{\text{rad}}(\tilde{I} \ominus R)$.

In order to verify the existance of such QESs and the leading order formula for the generalised entropy (3.1.10) we need to show that the variation of the generalised entropy around the putative extrema vanish.

The fact that $\partial\tilde{I} \subset \partial R$ means that there is a one-to-one map of the QESs ∂I to the endpoints ∂R of of R such that

$$\tilde{u}_a = u_i, \qquad v_a = u_i. \tag{3.6.5}$$

Here i denotes the image of a under the one-to-one map $\partial I \to \partial R$. Let us write the leading order correction as

$$\tilde{u}_a = u_i + \delta\tilde{u}_a, \qquad v_a = u_i + \delta v_a. \tag{3.6.6}$$

The corrections $\delta\tilde{u}_a$ and δv_a are not "infinitesimal" but are small compared with the spacing between the points in ∂R. We will see that the shift δv_a in the ingoing null coordinate of the QES is negative and equal to $-t_{\rm scr}$, where $t_{\rm scr}$ is the scrambling time.

The aim will be to prove the existence of a solution for $\delta\tilde{u}_a$ and δv_a, although the actual values will not be needed to determine the generalised entropy. We can rewrite (3.6.6) in Kruskal coordinates as

$$U_a = -\lambda_a U_i, \qquad V_a = -\mu_a/U_i, \tag{3.6.7}$$

where

$$\lambda_a = e^{2\pi T\delta\tilde{u}_a}, \qquad \mu_a = e^{2\pi T\delta v_a}. \tag{3.6.8}$$

Here the temperature is evaluated at u_i.

To proceed we make the ansatz that

$$\mu_a = \mathcal{O}(s), \qquad \lambda_a = \mathcal{O}(1). \tag{3.6.9}$$

We will then show that there exists a solution consistent with this ansatz. The ansatz implies that $U_aV_a = \mathcal{O}(s) \ll 1$ which means that the QESs lie close to the horizon. The ansatz for μ_a also implies that to leading order

$$\delta v_a \approx -\frac{1}{2\pi T(u_i)} \log \frac{N}{S_{\rm BH}(u_i) - S_0} = t_{\rm scr}. \tag{3.6.10}$$

The generalised entropy consists of two parts and we consider the variation of each term separately. Since the QESs lie close to the horizon we can use the near-horizon approximation for the area term

$$\frac{A}{4G} = |\partial I| S_0 + \sum_{a=1}^{|\partial I|} (S_{\rm BH}(v_a) - S_0)(1 - 2U_aV_a) + \dots \tag{3.6.11}$$

Varying with respect to $\tilde{u}_a$ gives

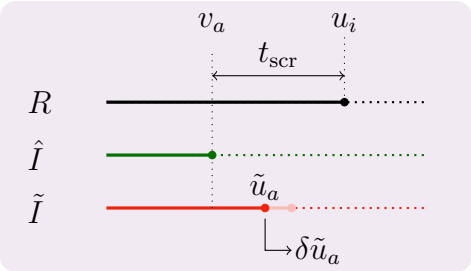
$$\frac{1}{4G}\frac{\partial A}{\partial\tilde{u}_a} = \frac{4\pi T\lambda_a\mu_a}{s} + \dots, \tag{3.6.12}$$

whilst varying with respect to v_a gives

$$\frac{1}{4G}\frac{\partial A}{\partial v_a} = -\frac{4\pi T \lambda_a \mu_a}{s} + \frac{\mathrm{d}S_{\mathrm{BH}}}{\mathrm{d}v_a} + \dots \tag{3.6.13}$$

To leading order we can replace $\tilde{u}_a$ and v_a with u_i in the arguments of the Bekenstein-Hawking entropy S_{BH} and the temperature T, as these are slowly varying in the adiabatic limit. In the following, any slowly varying function, such $T, S_{\mathrm{BH}}, S_{\mathrm{rad}}$, etc, is implicitly evaluated at u_i, unless stated otherwise.

The second contribution to the variation of the generalised entropy comes from the QFT entropy term $S_{\mathrm{QFT}}(I \cup R)$. We first consider the variation with respect to $\tilde{u}_a$. The effect on the modes collected in $I \cup R$ can be understood visually:



This variation leaves $\hat{I}$ unchanged and effectively converts modes originally contained only in R to modes contained in $\tilde{I} \cup R$, which leads to the gradient

$$-\frac{\mathrm{d}S^{\mathbb{T}}_{\mathrm{rad}}}{\mathrm{d}\tilde{u}_a} + \frac{\mathrm{d}S^{\mathbb{R}}_{\mathrm{rad}}}{\mathrm{d}\tilde{u}_a}. \tag{3.6.14}$$

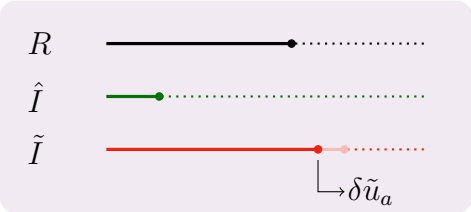
Hence

$$\begin{aligned}\frac{\partial}{\partial\tilde{u}_a}S_{\mathrm{QFT}}(I \cup R) &= \frac{\partial S_{\mathrm{Weyl}}(I)}{\partial\tilde{u}_a} - \frac{\mathrm{d}S^{\mathbb{T}}_{\mathrm{rad}}}{\mathrm{d}\tilde{u}_a} + \frac{\mathrm{d}S^{\mathbb{R}}_{\mathrm{rad}}}{\mathrm{d}\tilde{u}_a} \\ &= -\frac{\pi N T}{6} - \frac{\mathrm{d}S^{\mathbb{T}}_{\mathrm{rad}}}{\mathrm{d}\tilde{u}_a} + \frac{\mathrm{d}S^{\mathbb{R}}_{\mathrm{rad}}}{\mathrm{d}\tilde{u}_a},\end{aligned} \tag{3.6.15}$$

where the quantity $S_{\mathrm{Weyl}}(I)$ was defined in (3.5.28). Whilst this quantity provides a subleading contribution to the generalised entropy, its gradient is not. Since the first term in second line is just the gradient of the thermal entropy (3.1.5) we have

$$\frac{\partial}{\partial\tilde{u}_a}S_{\mathrm{QFT}}(I \cup R) = -\frac{\mathrm{d}S_{\mathrm{rad}}}{\mathrm{d}\tilde{u}_a} - \frac{\mathrm{d}S^{\mathbb{T}}_{\mathrm{rad}}}{\mathrm{d}\tilde{u}_a} + \frac{\mathrm{d}S^{\mathbb{R}}_{\mathrm{rad}}}{\mathrm{d}\tilde{u}_a}, \qquad \text{for } \tilde{u}_a < u_i. \tag{3.6.16}$$

For the case that $\tilde{u}_a > u_i$:



Following the same reasoning, the variation is

$$\frac{\partial}{\partial \tilde{u}_a} S_{\mathrm{QFT}}(I \cup R) = \frac{\partial S_{\mathrm{Weyl}}(I)}{\partial \tilde{u}_a} + \frac{\mathrm{d}S_{\mathrm{rad}}}{\mathrm{d}\tilde{u}_a} = 0, \qquad \text{for } \tilde{u}_a > u_i. \tag{3.6.17}$$

This analysis appears to imply that $S_{\mathrm{QFT}}(I \cup R)$ is discontinuous as we vary $\tilde{u}_a$ across u_i. This is an artefact of relying on the thermodynamic formulas for entropies which are not valid when $\tilde{u}_a$ and u_i are separated by less than the thermal scale. However, we do not need the explicit expression describing this smooth crossover in order to establish the existence of extrema. It is sufficient to write

$$\frac{\partial}{\partial \tilde{u}_a} S_{\mathrm{QFT}}(I \cup R) = \left[-\frac{\mathrm{d}S_{\mathrm{rad}}}{\mathrm{d}\tilde{u}_a} - \frac{\mathrm{d}S^{\mathbb{T}}_{\mathrm{rad}}}{\mathrm{d}\tilde{u}_a} + \frac{\mathrm{d}S^{\mathbb{R}}_{\mathrm{rad}}}{\mathrm{d}\tilde{u}_a} \right] f(\lambda_a), \tag{3.6.18}$$

for some smooth function f which varies between 0 and 1 as its argument varies between 1 and ∞, that is, f is a sort of smooth version of the Heaviside function.

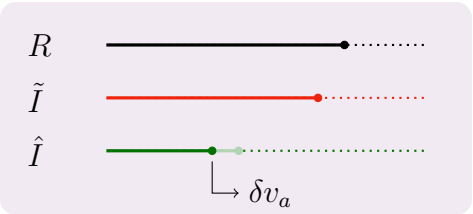
As a useful consistency check, in the case with a trivial greybody factor $\Gamma = 1$, where there are no reflected modes and the variation of $S_{\mathrm{QFT}}(I \cup R)$ can be computed using exact QFT techniques, we have (B.1.11)

$$\frac{\partial}{\partial \tilde{u}_a} S_{\mathrm{QFT}}(I \cup R) = \frac{-\pi N T}{3} \frac{U_a}{U_a - U_i} = -2 \frac{\mathrm{d}S_{\mathrm{rad}}}{\mathrm{d}\tilde{u}_a} \frac{\lambda_a}{\lambda_a + 1}. \tag{3.6.19}$$

This precisely matches (3.6.18) when $S^{\mathbb{R}}_{\mathrm{rad}} = 0$ and $S^{\mathbb{T}}_{\mathrm{rad}} = S_{\mathrm{rad}}$ if we identify

$$f(\lambda_a) = \frac{\lambda_a}{\lambda_a - 1}. \tag{3.6.20}$$

We now consider the variation of $S_{\mathrm{QFT}}(I \cup R)$ with respect to v_a. In this case:



In this case the variation leaves $\tilde{I}$ unchanged and effectively converts modes originally contained only in $\tilde{I} \cap R$ to modes contained in $\hat{I} \cup \tilde{I} \cup R$. This doesn't lead to a change in which leads to the gradient

$$-\frac{\mathrm{d}S^{\mathbb{R}}_{\mathrm{rad}}}{\mathrm{d}v_a}, \tag{3.6.21}$$

since the overstate for the three sets of modes is pure. Hence

$$\frac{\partial}{\partial v_a} S_{\mathrm{QFT}}(I \cup R) = \frac{\partial S_{\mathrm{Weyl}}(I)}{\partial v_a} - \frac{\mathrm{d}S^{\mathbb{R}}_{\mathrm{rad}}}{\mathrm{d}v_a} = \frac{\mathrm{d}S_{\mathrm{rad}}}{\mathrm{d}v_a} - \frac{\mathrm{d}S^{\mathbb{R}}_{\mathrm{rad}}}{\mathrm{d}v_a}. \tag{3.6.22}$$

Putting everything together gives us a pair of equation for λ_a and μ_a:

$$\begin{aligned}\frac{4\pi T\lambda_a\mu_a}{s}+\left[-\frac{\mathrm{d}S_{\mathrm{rad}}}{\mathrm{d}\tilde{u}_a}-\frac{\mathrm{d}S^{\mathbb{T}}_{\mathrm{rad}}}{\mathrm{d}\tilde{u}_a}+\frac{\mathrm{d}S^{\mathbb{R}}_{\mathrm{rad}}}{\mathrm{d}\tilde{u}_a}\right]f(\lambda_a)=0,\\ -\frac{4\pi T\lambda_a\mu_a}{s}+\frac{\mathrm{d}S_{\mathrm{BH}}}{\mathrm{d}v_a}+\frac{\mathrm{d}S_{\mathrm{rad}}}{\mathrm{d}v_a}-\frac{\mathrm{d}S^{\mathbb{R}}_{\mathrm{rad}}}{\mathrm{d}v_a}=0.\end{aligned} \tag{3.6.23}$$

The second equation determines the product $\lambda_a\mu_a$. Taking the sum of the two equations then determines λ_a as the solution of

$$f(\lambda_a)=\frac{\frac{\mathrm{d}S_{\mathrm{rad}}}{\mathrm{d}v_a}-\frac{\mathrm{d}S^{\mathbb{R}}_{\mathrm{rad}}}{\mathrm{d}v_a}+\frac{\mathrm{d}S_{\mathrm{BH}}}{\mathrm{d}v_a}}{\frac{\mathrm{d}S_{\mathrm{rad}}}{\mathrm{d}\tilde{u}_a}-\frac{\mathrm{d}S^{\mathbb{R}}_{\mathrm{rad}}}{\mathrm{d}\tilde{u}_a}+\frac{\mathrm{d}S^{\mathbb{T}}_{\mathrm{rad}}}{\mathrm{d}\tilde{u}_a}}. \tag{3.6.24}$$

To leading order all the quantities on the RHS are evaluated at u_i. A solution exists so long as the RHS is valued between 0 and 1. The denominator is positive by the Araki-Lieb inequality[10] and the fact that combined state of the transmitted and reflected modes is in a mixed state. Since $\mathrm{d}S^{\mathbb{T}}_{\mathrm{rad}}/\mathrm{d}\tilde{u}_a>0$ whilst $\mathrm{d}S_{\mathrm{BH}}/\mathrm{d}v_a<0$, a solution will exist so long as the numerator is nonnegative,

$$\frac{\mathrm{d}S_{\mathrm{rad}}}{\mathrm{d}v_a}-\frac{\mathrm{d}S^{\mathbb{R}}_{\mathrm{rad}}}{\mathrm{d}v_a}+\frac{\mathrm{d}S_{\mathrm{BH}}}{\mathrm{d}v_a}\geq 0. \tag{3.6.25}$$

This inequality is always satisfied, as we show in appendix B.2. The inequality is saturated in the limit where $\Gamma\to 0$ which corresponds to the limit where all the radiation is reflected back into the black hole. Notice that since all the quantities in (3.6.24) are evaluated at u_i, the parameter s does not appear. This means that the solution for λ_a is $\mathcal{O}(1)$, which is consistent with the ansatz (3.6.9). This completes the proof.

3.7 Page's model and irreversibility

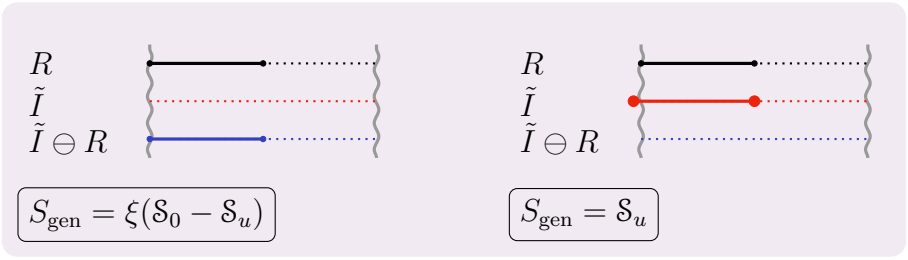
In this section we investigate the effect of thermodynamic irreversibility on the entropy of subsets of radiation. Generically there is a nontrivial competition between multiple islands, a feature that naively appears to be at odds with Page's model. In this section we review Page's random unitary model and then discuss a generalisation of the model which involves a nested sequence of random isometries, in order to incorporate thermodynamic irreversibility, and resolve the potential puzzle.

For simplicity, in this section we will assume that the extremal entropy S_0 is negligible and ignore its contribution. This restriction is not necessary and its effect can straightforwardly be reintroduced.

[10] This can also be proved directly for this case using the formula for the entropy flux in terms of the expectation value of the number operator (3.5.10) and the greybody factor Γ which takes values in the interval $[0,1]$.

3.7.1 Page curve

To derive the Page curve, we consider an interval $R = [0, u]$ which captures all the Hawking radiation emitted by the black hole from the time of formation $u = 0$ up to time u. There are two possible islands:



The first is the empty island $I = \emptyset$, the analogue of Hawking's result, which has generalised entropy

$$S_{\text{gen}}(R) = S^{\mathbb{T}}_{\text{rad}}([0, u]) = \xi(\mathcal{S}_0 - \mathcal{S}_u), \tag{3.7.1}$$

where $\mathcal{S}_u \coloneqq S_{\text{BH}}(u)$ and ξ is the greybody coefficient (3.7.18) which measures the ratio of the entropy fluxes of the transmitted radiation and the black hole. The second island has $I = \tilde{R}$, or equivalently $\tilde{I} = R$. Since $\tilde{I} \ominus R = \emptyset$, the generalised entropy with this island is

$$S_{\text{gen}}(I \cup R) = S_{\text{BH}}(u) = \mathcal{S}_u. \tag{3.7.2}$$

The entropy of R is then given by

$$S(R) = \min\{S^{\mathbb{T}}_{\text{rad}}([0, u]), S_{\text{BH}}(u)\}, \tag{3.7.3}$$

which is the Page curve. The Page time u_{Page} occurs when

$$\mathcal{S}_{u_{\text{Page}}} = \frac{\xi}{1 + \xi}\mathcal{S}_0. \tag{3.7.4}$$

In the reversible limit $\xi \to 1$, the Page time occurs when the black hole has lost exactly half its entropy. Irreversibility means that the Page time occurs significantly before the black hole has lost exactly half its entropy.

3.7.2 Page's model

In a chaotic quantum system the time evolution operator is a complicated unitary matrix for late times. It is sometimes useful to model it as a random unitary matrix, drawn from the Haar measure on the unitary group. Page's model is based on this idea. The state $|\Psi\rangle$ of the black hole, B, and its radiation, R, viewed as a bipartite quantum system with Hilbert space $\mathcal{H}_B \otimes \mathcal{H}_R$, is modelled as a Haar random state, that is, $|\Psi\rangle = U|\Psi_0\rangle$, where U is a Haar random unitary and $|\Psi_0\rangle$ is a fixed reference state

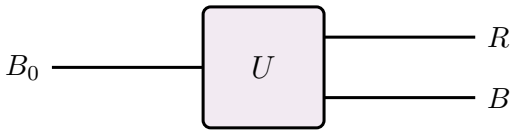


Figure 3.6: Page's model

(which can be thought of as the initial state of the black hole). Actually, for Page's model we won't need the full machinery of Haar random unitaries since $|\Psi\rangle$ just picks out a column of U which is equivalent to drawing a random vector uniformly from the unit sphere in Hilbert space. Equivalently, $|\Psi\rangle = |\psi\rangle \,/\, \||\psi\rangle\|$, where $|\psi\rangle$ is a vector whose components ψ_i are independent standard complex Gaussian random variables.

We will be interested in the (average) von Neumann entropy of the radiation subsystem. The von Neumann entropy $-\mathrm{Tr}\rho\log\rho$ can be extracted from the Rényi entropies, $\frac{1}{1-n}\log\mathrm{Tr}\rho^n$, by analytic continuation to $n=1$. The key idea of the replica trick is the observation that the n^{th} moment of ρ_R, the reduced state on the radiation subsystem, can be written as

$$\mathrm{Tr}\rho_R^n = \mathrm{Tr}[\tau_R\,\rho_R^{\otimes n}], \tag{3.7.5}$$

where on the RHS the trace is taken over the Hilbert space $\mathcal{H}_R^{\otimes n}$ and τ_R denotes the cyclic permutation operator on $\mathcal{H}_R^{\otimes n}$ which acts by cyclically permuting the n copies of $\mathcal{H}_R$.[11] The other crucial observation we will need is that for Gaussian random vectors[12]

$$\mathbb{E}[\psi^{\otimes n}] = \sum_{\pi\in S_n}\pi. \tag{3.7.6}$$

This follows from applying Wick's theorem to $\mathbb{E}[\psi_{i_1}\dots\psi_{i_n}\overline{\psi}_{j_1}\dots\overline{\psi}_{j_n}]$ since the ways we can pair the ψ_i with their complex conjugates $\overline{\psi}_i$ are determined by permutations $\pi\in S_n$. For our choice of Gaussian $|\psi\rangle$, the norm $\||\psi\rangle\|$ is independent of the normalised vector $|\psi\rangle \,/\, \||\psi\rangle\|$, so $\mathbb{E}[\Psi^{\otimes n}] = \mathbb{E}[\psi^{\otimes n}]/\mathbb{E}[\langle\psi|\psi\rangle^n]$. We may then compute the average moments

$$\begin{aligned}
\mathbb{E}\mathrm{Tr}\rho_R^n &= \mathbb{E}\mathrm{Tr}[\tau_R\,\rho_R^{\otimes n}] \\
&= \mathbb{E}\mathrm{Tr}[(\mathrm{id}_B\otimes\tau_R)\,\Psi^{\otimes n}] \\
&= \frac{\sum_{\pi\in S_n}\mathrm{Tr}[(\mathrm{id}_B\otimes\tau_R)\pi_{BR}]}{\sum_{\pi\in S_n}\mathrm{Tr}[\pi_{BR}]} \\
&= \frac{\sum_{\pi\in S_n} d_B^{\#(\pi)} d_R^{\#(\pi^{-1}\tau)}}{\sum_{\pi\in S_n} d^{\#(\pi)}}.
\end{aligned} \tag{3.7.7}$$

[11] For any permutation $\pi\in S_n$ we define the operator π (denoted by the same symbol) on $\mathcal{H}^{\otimes n}$ by its action on product states $\pi\,|\psi_1\rangle\otimes\cdots\otimes|\psi_n\rangle \coloneqq |\psi_{\pi^{-1}(1)}\rangle\otimes\cdots\otimes|\psi_{\pi^{-1}(n)}\rangle$ and then extend by linearity. Notice that it follows from the definition that π is unitary $\pi^\dagger=\pi^{-1}$.

[12] We adopt the convention that a vector $|\psi\rangle$ without the $|\ \ \rangle$ refers to the density matrix $\psi=|\psi\rangle\langle\psi|$.

Here $d_X = \dim \mathcal{H}_X$ and $\#(\pi)$ counts the number of disjoint cycles (including trivial cycles) in π. The term in the denominator can be computed exactly

$$\sum_{\pi \in S_n} d^{\#(\pi)} = d(d+1)\dots(d+n-1). \tag{3.7.8}$$

In the limit of large d, this may be approximated by d^n. This approximation may alternatively be derived by noting that $\#(\pi)$ is maximised for the identity permutation which consists of a product of trivial cycles, $\#(\mathrm{id}) = n$. Using the same logic, it is simple to determine the behaviour of the sum in the numerator when one of the subsystems is much larger than the other. The sum is dominated by the term $\pi = \mathrm{id}$ when $d_R \ll d_B$ and by the term $\pi = \tau$ when $d_R \gg d_B$,

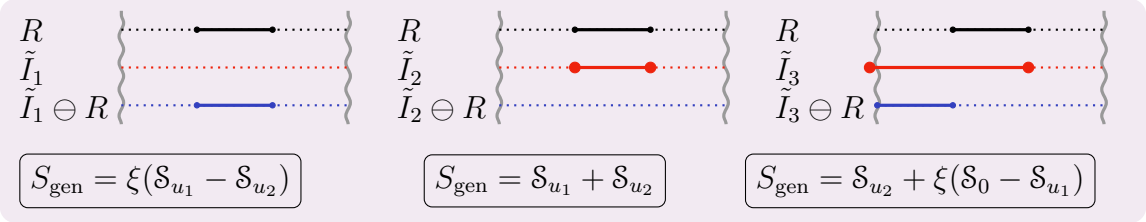
$$\mathbb{E}\mathrm{Tr}\rho_R^n = \begin{cases} d_R^{1-n} & \text{for } d_R \ll d_B, \\ d_B^{1-n} & \text{for } d_R \gg d_B. \end{cases} \tag{3.7.9}$$

Analytically continuing the Rényi entropy to $n = 1$ gives the Page curve (3.7.3), if we identify the dimension of the Hilbert space of the radiation and black hole subsystems as

$$\log d_R = S^{\mathbb{T}}_{\mathrm{rad}}([0,u]), \qquad \log d_B = S_{\mathrm{BH}}(u). \tag{3.7.10}$$

3.7.3 Single interval

Now consider the case that R no longer begins at $u = 0$ but rather $R = [u_1, u_2]$ with $u_1 > 0$. There are three possible islands:



The entropy is given by

$$S(R) = \min\left\{(S^{\mathbb{T}}_{\mathrm{rad}}([u_1,u_2]), S_{\mathrm{BH}}(u_1) + S_{\mathrm{BH}}(u_2), S^{\mathbb{T}}_{\mathrm{rad}}([0,u_1]) + S_{\mathrm{BH}}(u_2)\right\}. \tag{3.7.11}$$

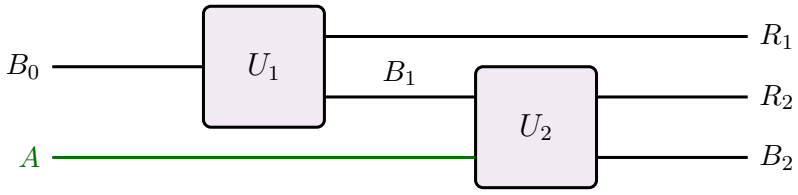
The second island I_2 can only contribute if $S^{\mathbb{T}}_{\mathrm{rad}}([0,u_1]) < S_{\mathrm{BH}}(u_1)$ or equivalently if $\xi > 1$, that is, if evaporation is irreversible. In the reversible limit, $\xi \to 1$, there is only a competition between the empty island $I_1 = \emptyset$ and the island I_3, and the relation to Page's result is immediate. In fact, for a generic subset R, one can show that there will only ever be two competing islands in the reversible limit: the first is the empty island, and the second is the island with $\tilde{I} = R \cup R'$, where R' is the complement of R in the radiation. In this case the entropy is given by

$$S(R) = \min\left\{S^{\mathbb{T}}_{\mathrm{rad}}(R), S_{\mathrm{BH}}(u) + S^{\mathbb{T}}_{\mathrm{rad}}(R')\right\}, \tag{3.7.12}$$

in line with Page's result.

3.7.4 Random isometry model

The reason Page's result doesn't always apply is because the model misses the fact that as the number of qubits describing the black hole (the Bekenstein-Hawking entropy) decreases over time, the total number of qubits describing the black hole and its radiation (the total thermodynamic entropy) increases. Only in the reversible limit does the total number of qubits remain constant. To account for this we can add some ancilla qubits. The simplest example of the model is:



In this model the time evolution is modelled as a nested sequence of Haar random unitaries with some ancilla qubits A, in some fixed state, added to model thermodynamic irreversibility. Equivalently, we can think of the random unitary U_2 acting on the fixed ancilla qubits A as defining a random isometry $\mathcal{H}_{B_1} \to \mathcal{H}_{R_2} \otimes \mathcal{H}_{B_2}$. Shortly, we will connect this model to the example of a single interval studied in the previous section.

In contrast to Page's model, we will need to work with the full machinery of Haar random unitaries. We will explain this shortly; for now we simply state the result for the model above. When the dimensions of the subsystems are large, averaging over the unitaries U_1 and U_2 gives a double sum over the permutation group:

$$\mathbb{E}\mathrm{Tr}\rho_R^n \approx \sum_{\pi_1,\pi_2\in S_n} d_{R_1}^{-d(\pi_1,\mathrm{id})} d_{B_1}^{-d(\pi_1,\pi_2)} d_{R_2}^{-d(\pi_2,\tau)} d_{B_2}^{-d(\pi_2,\mathrm{id})}, \tag{3.7.13}$$

where $d(\pi_1,\pi_2) = n - \#(\pi_1^{-1}\pi_2)$ is the Cayley distance function on S_n which counts the minimal of transpositions needed to go from π_1 to π_2. In the limit where the dimensions of the subsystems are large, only the identity permutation id and cyclic permutation τ contribute to the sum (see appendix B.3). The dominant contributions are:

$$\mathbb{E}\mathrm{Tr}\rho_R^n \supset d_{R_2}^{1-n} + (d_{B_1}d_{B_2})^{1-n} + (d_B d_{R_1})^{1-n}, \tag{3.7.14}$$

which correspond to the terms where (π_1,π_2) is equal to $(\mathrm{id},\mathrm{id}), (\mathrm{id},\tau)$, and (τ,τ). Analytically continuing the Rényi entropy to $n=1$ gives the entropy

$$\mathbb{E}S(R) = \min\{\log d_{R_2}, \log d_{B_1} + \log d_{B_2}, \log d_{B_2} + \log d_{R_2}\}. \tag{3.7.15}$$

To connect this with the example of a single interval studied in the previous section we identify the radiation subsystems R_1 with $R' = [0,u_1]$ and R_2 with $R = [u_1,u_2]$ as well as

$$\log d_{R_1} = S^{\mathbb{T}}_{\mathrm{rad}}([0,u_1]), \qquad \log d_{R_2} = S^{\mathbb{T}}_{\mathrm{rad}}([u_1,u_2]), \qquad \log d_{B_2} = S_{\mathrm{BH}}(u_2). \tag{3.7.16}$$

With this identification (3.7.15) agrees perfectly with the competition between the three islands (3.7.11). Notice that the role of the ancilla qubits is simply to allow $d_{R_2}d_{B_2} > d_{B_1}$, that is,

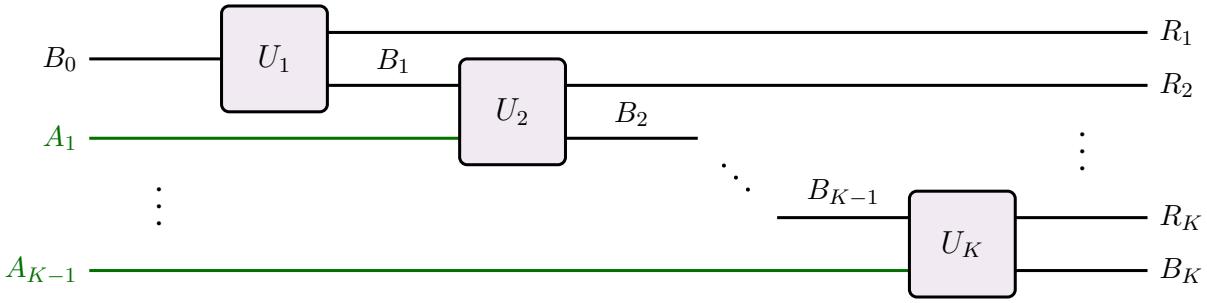
$$S^{\mathbb{T}}_{\text{rad}}([u_1, u_2]) > S_{\text{BH}}(u_1) - S_{\text{BH}}(u_2). \tag{3.7.17}$$

In particular we can make the identification

$$\xi = 1 + \frac{\log d_A}{\log d_{B_1}/d_{B_2}} \geq 1. \tag{3.7.18}$$

The inequality is saturated in the reversible limit when $d_A \to 1$ and the isometry $\mathcal{H}_{B_1} \to \mathcal{H}_{R_2} \otimes \mathcal{H}_{B_2}$ is unitary.

In general the random isometry model with K time steps, each chosen to be at least a scrambling time, is:



Let R denote a subset of $R_1 \cup \ldots R_K \cup B_K$ and $\bar{R}$ denote its complement. Using the Replica trick the n^{th} moment of the ρ_R can be written as

$$\mathbb{E}\text{Tr}\rho_R^n = \mathbb{E}\text{Tr}[\tau_R\, \rho_R^{\otimes n}] = \mathbb{E}\text{Tr}[(\text{id}_{\bar{R}} \otimes \tau_R)\, \Psi^{\otimes n}]. \tag{3.7.19}$$

Here $|\Psi\rangle = V_K \ldots V_1 |\Psi_0\rangle$, where $V_k = U_k |0\rangle_{A_{k-1}}$ and $|0\rangle_{A_{k-1}}$ is some fixed state on the anicilla. It's convenient to use cyclicity of the trace to write

$$\mathbb{E}\text{Tr}[(\text{id}_{\bar{R}} \otimes \tau_R)\, \Psi^{\otimes n}] = \mathbb{E}\text{Tr}[(V_1^\dagger \ldots V_K^\dagger)^{\otimes n}(\text{id}_{\bar{R}} \otimes \tau_R)(V_K \ldots V_1)^{\otimes n}\, \Psi_0^{\otimes n}]. \tag{3.7.20}$$

The average can be computed in the following way. Let $U : \mathcal{H}_X \to \mathcal{H}_Y$ be a Haar random unitary and Y an operator on $\mathcal{H}_Y^{\otimes n}$. Since the Haar measure is right-invariant,

$$\mathbb{E}[(U^\dagger)^{\otimes n} \mathsf{Y} U^{\otimes n}] \tag{3.7.21}$$

commutes with the diagonal action of the unitary group. A consequence of Schur-Weyl duality is that any such operator must be a linear combination of permutation operators. Hence,

$$\mathbb{E}[(U^\dagger)^{\otimes n} \mathsf{Y} U^{\otimes n}] = \sum_{\pi \in S_n} W_\pi(\mathsf{Y}) \pi_X, \tag{3.7.22}$$

for some linear functionals W_π. The exact formula for W_π can be expressed in terms of the Weingarten function for the unitary group and will not be important for us. For our purposes, the asymptotic formula,

$$W_\pi(\mathsf{Y}) \approx \frac{1}{d_Y^n}\mathrm{Tr}[\pi^{-1}\mathsf{Y}], \tag{3.7.23}$$

which is valid when the dimension of Hilbert is large, will suffice. Notice that if $\mathsf{Y} = \pi'$ is a permutation operator then this becomes

$$W_\pi(\pi') \approx d_Y^{-d(\pi,\pi')}. \tag{3.7.24}$$

Using these facts, we find that the n^{th} moment of ρ_R is given by a nested sum over K copies of the symmetric group S_n,

$$\mathbb{E}\mathrm{Tr}\rho_R^n \approx \prod_{k=1}^{K}\sum_{\pi_k\in S_n} d_{B_k}^{-d(\pi_k,\pi_{k+1})} d_{R_k}^{-d(\pi_k,\sigma_k)}, \tag{3.7.25}$$

where σ_k is the cyclic permutation τ if $R_k \subset R$, and as the identity permutation id otherwise. Similarly, π_{K+1} is the cyclic permutation τ if $B_K \subset R$, and the identity permutation id otherwise. In the limit where the dimensions of the subsystems are large, only the identity permutation id and cyclic permutation τ contribute to each of the K sums (see appendix B.3). In fact, not all of the 2^K combinations can contribute, as we saw in (3.7.14) in the example of the model with two time steps. There are effectively selection rules between an adjacent pair of permutations (π_k, π_{k+1}): if $R_k \subset R$ then (τ, id) will not contribute, whilst if $R_k \nsubseteq R$ then (id, τ) will not contribute. These selection rules equivalent are equivalent to the claim that in the island in the stream formula the points in ∂I must correspond to the points in ∂R.

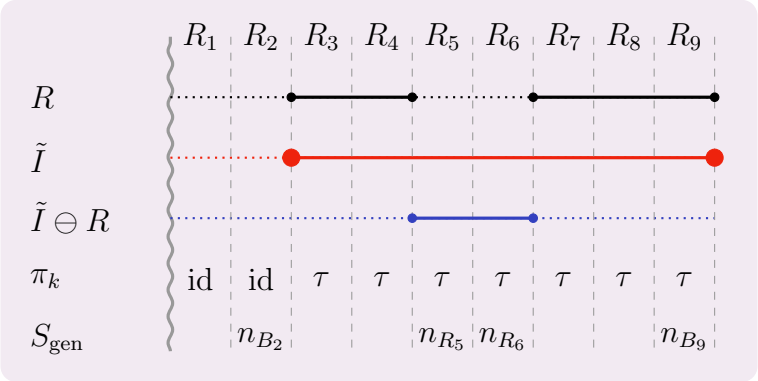


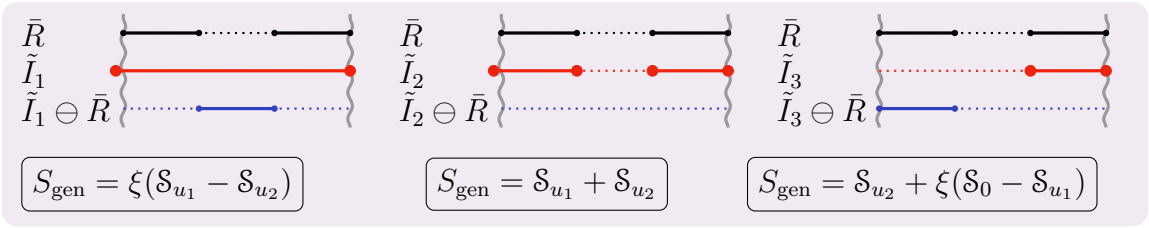
Figure 3.7: An example illustrating the connection between certain permutations π_k which contribute to the average Rényi entropy and island. Here we adopted the shorthand $n_{B_k} = \log d_{B_k} = S_{\mathsf{BH}}(u_k)$ and $n_{R_k} = \log d_{R_k} = S^{\mathbb{T}}_{\mathsf{rad}}([u_{k-1}, u_k])$. For this example, $S_{\mathsf{gen}}(I \cup R) = n_{B_2} + n_{R_5} + n_{R_6} + n_{B_9}$.

To illustrate this, it is instructive to consider the example shown in figure 3.7 where the relation between the permutations which can dominate the sum and the corresponding island contribution is clear, namely,

$$\tilde{I} = \bigsqcup_{\substack{k: \\ \pi_k = \tau}} R_k. \tag{3.7.26}$$

3.7.5 Decoupling

Another simple application of the island in the stream recipe is to study when there are no correlations between two subsets of radiation R_1 and R_2. This happens when the mutual information $I(R_1 : R_2)$ vanishes, or equivalently, when the reduced state on $R_1 \cup R_2$ factorises, that is $\rho_{R_1 R_2} = \rho_{R_1} \otimes \rho_{R_2}$. To investigate this, consider the setup of section 3.7.3 and let $\bar{R}$ denote the complement of R. The relevant contributions to the entropy of $\bar{R}$ are:



By comparing with the discussion in section 3.7.3 we can see that $S(R) = S(\bar{R})$, as required by unitarity. Turning to the issue of decoupling, notice that when the island I_2 or I_3 contribute, the entropy of $\bar{R}$ factorises: $S(\bar{R}) = S(B) + S(R')$ and hence $I(B : R') = 0$. Here R' denotes the complement of R in the radiation, that is $\bar{R} = R' \cup B$. When this is the case, R also receives a contribution from a nonempty island. This motivates the following rule: there is no correlation between two subsets of radiation R_1 and R_2 which are separated by a large enough region R such that R receives a contribution from a nonempty island.[13]

3.8 Conclusion and outlook

We have see how combining the fact that black holes evaporate very slowly in the semiclassical limit and thermodynamic limit with the QES prescription leads to a simple recipe for determining a class of islands I which can contribute to an arbitrary subset $R = \bigsqcup_k R_k$ of the radiation. The characteristic feature of this class of islands

[13] This is related to theorem IV.2 in [86].

was that their mirror endpoints corresponded to endpoints of the R_k. This feature, together with the entanglement structure of modes across the horizon allowed us to give a simple visual way of computing the generalised entropy which involved the symmetric difference $\tilde{I} \ominus R$.

These formulas were derived under the assumption that the matter sector consisted of N massless Dirac fermions. This choice was due to the fact that Dirac fermions are the only CFT for which the multi-interval von Neumann entropy is known analytically. It would be interested to try to extend these arguments to a generic CFT. This might be possible in the thermodynamic limit by making an ansatz that the QFT entropy term can be approximated by the leading term in OPE expansion for the twist fields in the replica geometry, as mentioned briefly in 3.2.4. This suspicion arises from its apparent success in a slightly different setup described in chapter 5.

We also showed how the islands in the stream formula could be generalised to account for a greybody factor, which in the JT model was implemented by changing the boundary conditions for matter fields along the AdS-Minkowski interface. In the limit where the scales of interest were much larger than the scrambling time, the generalisation was straightforward: the thermal entropy $S_{\text{rad}}(I \ominus R)$ was simply replaced by the entropy $S^{\mathbb{T}}_{\text{rad}}(I \ominus R)$ of the transmitted radiation. For simplicity, we worked under the assumption that the greybody factor $\Gamma(\omega)$ only depended on the frequency ω and the temperature T. With this assumption the effect of the greybody factor was simply to rescale the energy and entropy fluxes by a constant. It would be interesting to understand to what extent the islands in the stream formula holds if we relax the condition on the greybody factor.

The parameter ξ, which measured the ratio of the entropy fluxes of the transmitted radiation and the black hole, played an important role. Whilst generically there can be multiple possible islands which contribute to the entropy of the radiation, in the reversible limit $\xi \to 1$ we found that there were only two possible islands. Remarkably, this behaviour was captured by a generalisation of Page's model, where time evolution was implemented by a nested sequences of random unitary transformations. Calculations in this model were facilitated by averaging over the unitary group. Actually, since the Rényi entropies are self averaging, the Rényi entropy of a random state is, with extremely high probability, given by its average. Though, the fact that averaging appears is also not completely surprising since there are arguments that the gravitational path integral is actually computing a coarse grained quantity, where we average over some microscopic information [87]. This is motivated partly by the fact that pure JT gravity is genuinely dual to an ensemble of random Hamiltonians [35].

Although our results enable one to study the correlation between subsets of the radiation in the form of mutual information, they do not provide insight into the multipartite entanglement structure of the radiation. Exploring measures of multipartite entanglement, like negativity, which already exhibits a rich structure in random tensor networks [88], could provide valuable intuition in understanding this problem.

Another natural question to ask is how the islands in the stream formalism can be used to study the problem of information recovery for an object which has been thrown into a black hole. This the problem is the focus of the next chapter.

Chapter 4

Information recovery

This chapter primarily consists of a paper [59] written in collaboration with Zsolt Gyongyosi, Timothy J. Hollowood, S. Prem Kumar, and Andrea Legramandi. Section 4.5 contains unpublished work.

4.1 Introduction

In the previous chapters we have seen how the QES prescription provides the necessary accounting trick to obtain the Page curve. In this chapter we focus on a different aspect of the information problem, first studied by Hayden and Preskill [47]. Namely, suppose we throw a *diary*, that is, some quantum information, into a black hole. When can the information in the diary be recovered in the radiation?[1]

Using a similar random unitary model to Page [5], Hayden and Preskill showed that an old black hole behaves as a mirror, effectively reflecting back information thrown into it after the Page time. An understanding of how this result follows directly from a bulk argument was first given in [9, 10], using entanglement wedge reconstruction (EWR). The result follows because for an old black hole, that is, past the Page time, the entanglement wedge of the radiation R includes an island I which lies inside the horizon. The endpoint of the island, that is, the QES, lags behind the endpoint of R by a scrambling time, as shown in figure 4.1. Hence, by EWR, any information thrown into the black hole at least a scrambling time in the past is recoverable in the radiation.

The Hayden-Preskill decoding criterion was later refined by Hayden and Penington [82] who showed that the dimension of the code space, that is, the diary, affects the time scale for the information to be recoverable in the radiation. In particular, even

[1] More specially Hayden and Preskill asked when the recovery operation becomes information-theoretically possible. A related but distinct question is how difficult the recovery operation actually is [89]—a topic that will be discussed in section 4.4.

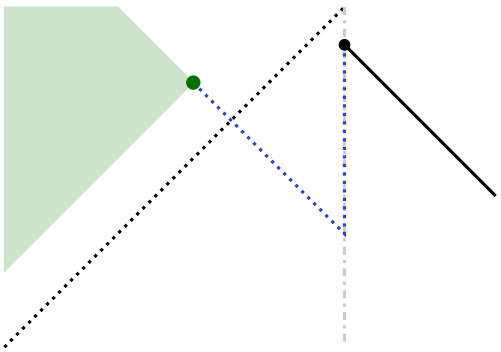


Figure 4.1: The relationship between the QES, shown in green, for an old black hole and the endpoint of R, shown in black. The entanglement wedge of the radiation includes a portion of the interior of the black hole (shaded green). The ingoing null coordinate v of the QES lags behind the outgoing null coordinate u of the endpoint of R by a scrambling time.

long after the Page time information recovery is only possible for sufficiently small code spaces. This is in stark contrast with the naive conclusion one would reach by assuming EWR to be exact, where being able to decode all small code subspaces automatically implies the ability to decode arbitrary subspaces. The result of Hayden and Penington is thus inextricably tied to the fact that EWR is merely approximate, rather than exact[2]. To decode larger code spaces one needs the entanglement wedge of the radiation to contain the diary for all states in the codes space, including mixed states. It turns out that this refined decoding criterion also follows directly from a bulk argument [10]. A mixed state on the code space essentially corresponds to diaries which carry entropy. By Bekenstein's bound [90], or rather its refinement due to Casini [91], a diary carrying entropy into the black hole necessarily carries energy and so backreacts on the geometry. Understanding the effect of this backreaction is crucial for understanding this refined version of the decoding criterion.

The aim of this work is to consider this problem in a model of JT gravity coupled to an auxiliary reservoir with the matter sector taken to be a CFT. In this setup we model the diary as an ingoing shockwave created by a local quench in the CFT, as described in section 2.5. This model is rich enough to capture the dynamics of the s-wave sector of the near-horizon limit of a near-extremal black hole in higher dimensions, yet simple enough that the backreaction of the diary can be solved exactly. This is because in JT gravity the metric is fixed to be a portion of AdS_2 and the dynamics is entirely encoded in the boundary action

$$I_{\text{JT}} = -\frac{\phi_r}{8\pi G}\int \mathrm{d}t\,\{F,t\}, \tag{4.1.1}$$

for the function F which relates the boundary time coordinate t, which is the natural time coordinate for an asymptotic observer, and the Poincaré time coordinate.

[2]The errors in reconstruction will be discussed explicitly in a simple example in section 4.5.

For a diary thrown into an old black hole one finds that the QES which is responsible for the downward portion of the Page curve gets stuck to the past of the diary, as shown in figure 4.2. At later times, a new QES which lies close to the horizon and to the future of the diary becomes relevant. When this new QES has minimal generalised entropy the diary lies in the entanglement wedge of the radiation. The delay in the transition to this new QES is what is responsible for reproducing the refined decoding criterion. Aspects of this problem for black holes in general dimensions were first considered in [10]. By working in a simple model, and with a concrete model for the diary, we are able to solve the backreaction problem exactly, in contrast to [10], where the backreaction problem could only be solved approximately, however, we reach the same conclusions.

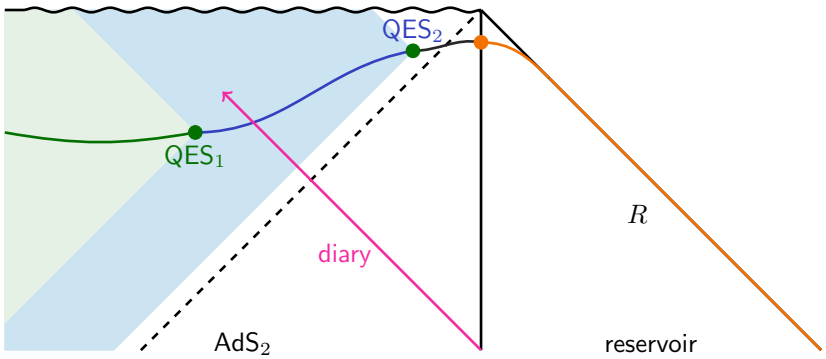


Figure 4.2: When a diary is thrown in from the boundary after the Page time, QES$_1$, which is responsible for the downward portion of the Page curve, gets stuck behind the diary. At later times, when QES$_2$ is the minimal QES, the diary lies in the entanglement wedge of the radiation.

One of the main contributions of this work is to study a variant of the decoding problem where we only have access to a subset of the radiation. In particular, we investigate whether the information in the diary can be recovered if we only have access to the late radiation, collected after the diary was thrown in. This variant of the decoding problem can be understood as testing the robustness of the encoding of the diary in the radiation against erasures, that is, loosing access to part of the radiation.

The argument for information recovery based on EWR is somewhat indirect, it involves taking entropies, which can computed directly from the gravitational path integral using the replica trick, and then applying the machinery of quantum information to it. A more direct argument in a toy model of JT gravity, the PSSY model, was given [11] using the Petz map. They asked whether an operator behind the horizon could be reconstructed as an operator on the radiation alone. This is nothing but the problem of information recovery stated in the Heisenberg picture rather than the Schrödinger picture. A theorem by Dénes Petz [92] in the theory of recovery channels provides a reconstruction which works whenever reconstruction is possible. The authors [11] then showed how the matrix elements of this reconstruction, the Petz map, could be computed directly from the gravitational path integral using the replica trick. Another

aim of this chapter is to consider a variant of this calculation to a different toy model, closely related to the random isometry model introduced in section 3.7.4. In this setting, we will show how to recover the refined decoding criterion of [82] as well as to understand salient features of irreversibility which are present in the JT model.

The plan of this chapter is as follows. We begin section 4.2 by reviewing some relevant aspects of the islands in the stream formula before discussing its generalisation to the case with a diary, with the detailed derivation deferred to appendix C.1. In section 4.3 carry out a more refined analysis where we solve exactly for the backreaction of the diary on the geometry. In section 4.4 we estimate the complexity of decoding a diary based on the Python's lunch conjecture [93]. In section 4.5 we show how to obtain the decoding criterion more directly in a toy model by considering reconstruction using the Petz map.

4.2 Islands in the stream with diaries

In this section we show how the island in the stream recipe generalises to the case where a diary is thrown into a black hole.

4.2.1 Review of islands in the stream without diaries

We begin by reviewing some of the relevant aspects of the usual island in the stream recipe before generalising to the case with a diary. In the semiclassical limit black holes evaporate very slowly. This corresponds to a sort of adiabatic limit where it is meaningful to associate a slowly varying temperature $T(t)$ to the black hole. In this limit, the QFT entropy of a subset $R = \bigsqcup_k R_k$ of the radiation is given by the thermodynamic formula

$$S_{\text{rad}}(R) = \frac{\pi c}{6} \int_R T(u) \mathrm{d}u, \tag{4.2.1}$$

when each of the intervals R_k and the spaces between them are much larger[3] than the thermal scale. Here the integral over R denotes integral over the range of the outgoing null coordinates which span R. That is, if the range of outgoing null coordinates u over the interval R_k is $[u_k, u'_k]$, then

$$S_{\text{rad}}(R_k) = \frac{\pi c}{6} \int_{u_k}^{u'_k} T(u) \mathrm{d}u. \tag{4.2.2}$$

Combining the adiabatic and thermodynamic limit with the QES formula leads to a very simple recipe for determining a class of islands I which can contribute to the entropy of a subset R of the radiation:

[3] With respect to the measure $T(u)\mathrm{d}u$ or $\frac{\mathrm{d}U}{|U|}$.

1. Under the reflection across the horizon $U \mapsto -U$ the outgoing null coordinates $\tilde{u}_{\partial I}$ of the QESs are in one-to-one correspondence with points in ∂R. The ingoing null coordinate $v_{\partial I}$ of the QESs lag behind by a scrambling time,

$$v_{\partial I} = \tilde{u}_{\partial I} - t_{\rm scr}, \qquad t_{\rm scr} = \frac{1}{2\pi T} \log \frac{S_{\rm BH} - S_0}{c}. \tag{4.2.3}$$

2. The generalised entropy of $I \cup R$ is given by

$$S_{\rm gen}(I \cup R) = \sum_{\substack{\text{disconnected} \\ \text{components} \\ \partial \tilde{I}_k \text{ of } \partial \tilde{I}}} S_{\rm BH}(u_{\partial \tilde{I}_k}) + S_{\rm rad}(\tilde{I} \ominus R), \tag{4.2.4}$$

   where $\tilde{I}$ is the reflection of I across the horizon. Here $\ominus$ denotes the symmetric difference, $\tilde{I} \ominus R = \tilde{I} \cup R - \tilde{I} \cap R$. Finally the von Neumann entropy $S(R) = \min_I S_{\rm gen}(I \cup R)$ of R is determined by minimising over possible islands.

In this section we will assume that the scrambling time $t_{\rm scr}$ is a subleading time scale. In section 4.3 we will keep track of its effect.

A feature of black holes in higher dimensions is that the spectrum of Hawking radiation is not actually thermal. Instead, it is modified by greybody factors. In the JT model one can mimic the effect of a greybody factor by modifying the boundary conditions along the AdS-Minkowski interface. For the case that the greybody factor $\Gamma(\omega)$ is only a function of the frequency ω and the temperature T, the greybody factor leads to a simple multiplicative relation between entropy flux $\mathrm{d}S^{\mathbb{T}}/\mathrm{d}u$ of the transmitted radiation $S^{\mathbb{T}}$ and the entropy flux of the Bekenstein-Hawking entropy,

$$\frac{\mathrm{d}S^{\mathbb{T}}}{\mathrm{d}u} = -\xi \frac{\mathrm{d}S_{\rm BH}}{\mathrm{d}u}. \tag{4.2.5}$$

The greybody coefficient $1 \leq \xi \leq 2$ characterises the irreversibility of the evaporation, with the limit $\xi \to 1$ corresponding to the reversible case. The generalisation of the islands in the stream formula (4.2.4) to the case with such a greybody mimicked in the JT model is particularly simple,

$$S_{\rm gen}(I \cup R) = \sum_{\substack{\text{disconnected} \\ \text{components} \\ \partial \tilde{I}_k \text{ of } \partial \tilde{I}}} S_{\rm BH}(u_{\partial \tilde{I}_k}) + S^{\mathbb{T}}_{\rm rad}(\tilde{I} \ominus R). \tag{4.2.6}$$

The only difference, as compared with (4.2.4), is that we have replace the thermal entropy $S_{\rm rad}$ of the radiation with the entropy of the transmitted radiation $S^{\mathbb{T}}_{\rm rad}$.

4.2.2 Adding a diary

It's possible to generalise the islands in the stream recipe to the case where we throw in a diary. We assume the diary is modelled as an ingoing shockwave which carries a constant energy E_D and entropy S_D. This will be explained in more detail in section 4.3.2.

A diary which carries entropy necessarily caries energy, and so backreacts on the geometry. This is simply a consequence of Bekenstein's bound[90, 91]. Alternatively, this follows from the first law of thermodynamics together with the generalised second law (GSL) [2] which requires

$$\Delta S_{\text{BH}} \geq S_D. \tag{4.2.7}$$

Here ΔS_{BH} is the jump in the Bekenstein-Hawking entropy as the diary is thrown in at time t_D. For simplicity, we will restrict to the case that the jump in the Bekenstein-Hawking entropy is small compared to the initial Bekenstein-Hawking entropy, that is,

$$\Delta S_{\text{BH}} \ll S_{\text{BH}}(t_D), \tag{4.2.8}$$

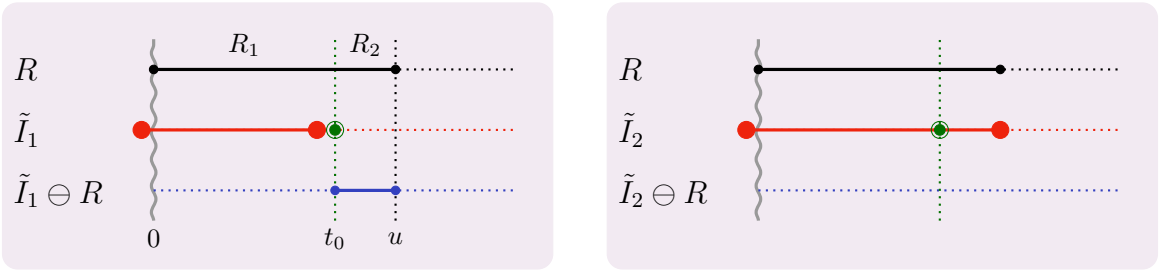
yet large enough to have an interesting effect. Large enough turns out to mean larger than the entropy of radiation emitted in a thermal time scale, that is, $\Delta S_{\text{BH}} \gtrsim c$.

We now describe how the islands in the stream recipe generalises to account for a diary. The detailed proof is relegated to appendix C.1. Without a diary, the outgoing null coordinates $\tilde{u}$ of the QESs are in one-to-one correspondence with points in ∂R. With a diary, the Bekenstein-Hawking entropy must now include the backreaction of the diary, in particular, it jumps discontinuously across $v = t_D$. In addition, we find a new QES which lies just to the past of the diary and, to leading order, has ingoing and outgoing null coordinates v and $\tilde{u}$ equal to t_D. A more refined analysis in section 4.3 pins down the location of this QES more precisely. The upshot is that in the presence of a diary is that outgoing null coordinates $\tilde{u}$ are now in one-to-one correspondence with $\partial R \cup \{t_D\}$, where $\{t_D\}$ represents the possibility of the new QES.

4.2.3 Information recovery in the early radiation

The first scenario we consider is the one considered by Hayden and Preskill [47], where the all the radiation emitted up to time u since the formation of the black hole at $u = 0$ is collected, that is, $R = [0, u]$. The question is: what is the minimum time needed to recover the information of the diary in R?

Note that we must have at least $u > t_0$. Consequently it's convenient to split R into two subsets: the early radiation, $R_1 = [0, t_D]$, and the late radiation, $R_2 = [t_D, u]$. The entropy of R involves a competition between three islands, of which two are nonempty:



In these diagrams the red dot on the left represents a QESs which lies at the point just before the black hole was formed by an infalling shockwave and contributes entropy S_0, which we are assuming is negligible. The green dot represents the diary. From the diagrams, it's straightforward to read off entropy

$$S(R) = \min\left\{S_{\text{rad}}(R), S_{\text{BH}}(t_D) + S_{\text{rad}}(R_2), S_{\text{BH}}(u) + S_D\right\}. \tag{4.2.9}$$

By EWR, the diary is recoverable in R when I_2 has minimal generalised entropy, that is,

$$S_{\text{BH}}(u) + S(D) < \min\left\{S_{\text{rad}}(R), S_{\text{BH}}(t_D) + S_{\text{rad}}(R_2)\right\}. \tag{4.2.10}$$

This means that if the diary is thrown in before the Page time then there is a direct transition between the empty island $\emptyset$ and I_2, so the diary is recoverable after an S_D's worth of additional radiation has been emitted after the Page time. On the other hand, if the diary is thrown in after the Page time there are two transitions: $\emptyset \to I_1 \to I_2$. In this case, the diary is recoverable when

$$S_{\text{rad}}(R_2) = \frac{\xi}{\xi + 1}\left(\Delta S_{\text{BH}} + S_D\right). \tag{4.2.11}$$

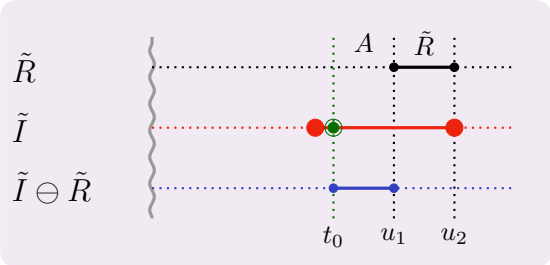
In obtaining this formula we have used (4.2.5). This delay between the time when the diary is thrown in and recovery time can be interpreted as the time it takes for the black hole to "process" the diary and return its information though Hawking radiation. In the reversible limit $\Delta S_{\text{BH}} = S_D$ and $\xi \to 1$, so the diary is recoverable when

$$S_{\text{rad}}(R_2) = S_D. \tag{4.2.12}$$

Hence, even long after the Page time reconstruction is only possible for sufficiently small diaries or code spaces. This state dependence is in line with expectations [47, 82, 94]. Notice that even though the early radiation never actually interacts with the diary, it can be useful to decoding. This is a characteristic feature of quantum error correcting codes, decoding is aided by having access to subsystems that never interacted with the encoded subsystem but which are entangled with subsystems that did.

4.2.4 Information recovery in the late radiation

In this section we consider the possibility of information recovery in the late radiation, assuming we don't have access to the early radiation. For an interval of late radiation $\tilde{R} = [u_1, u_2]$ with $u_1 > t_D$, recovery is possible the island I with $\tilde{I} = [t_D, u_2]$ has minimal generalised entropy:



It follows that recovery is always possible for a suitably small A and large enough $\tilde{R}$. For example, if $\tilde{R} = [t_D, u]$ is all the late radiation, then recovery is possible when

$$S_{\text{rad}}(\tilde{R}) \geq S_{\text{BH}}(u) + S_{\text{BH}}(t_D^-) + S_D. \tag{4.2.13}$$

This is sort of similar to the scenario in the previous section where we considered the possibility of a diary thrown in before the Page time. In this scenario we see that there is a further delay by $S_{\text{BH}}(t_D^-)$ due to the fact that we didn't have access to the early radiation.

4.3 Refined decoding criterion in JT gravity

In this section we carry out a more refined analysis where we solve for the backreaction of the diary on the geometry as it falls in. In JT gravity, where the metric is fixed to be a potion of AdS_2, this amount to solving for the function F which relates the boundary time coordinate t, which is the natural time coordinate for an asymptotic observer, and the Poincaré time coordinate. We begin by reviewing the setup without a diary. After adding a diary we solve for the new QES.

4.3.1 Review

The metric and dilaton profile can be written in Kruskal coordinates as

$$\mathrm{d}s^2 = -\frac{4\mathrm{d}U\mathrm{d}V}{(1+UV)^2}, \qquad \phi = \phi_r \left(\frac{F''(v)}{F'(v)} - \frac{1}{2}\frac{UF'(v)}{1+UV} \right), \tag{4.3.1}$$

where the function F which relates the Kruskal null coordinates U and V to the null coordinates u and v adapted to an asymptotic observer,

$$U = -1/F(u), \qquad V = F(v). \tag{4.3.2}$$

An evaporating black hole can be formed by heating an zero-temperature black hole and then letting it evaporate into the reservoir, as described in section 2.5. The exact solution for the function F can be expressed in terms of modified Bessel functions (2.5.4), however, there is an simple yet excellent WKB like solution (2.5.10) which is valid until very late times,

$$F(t) \approx \exp\left(2\pi \int_0^t \mathrm{d}t'\, T(t')\right), \qquad T(t) = T_0 e^{-kt/2}. \tag{4.3.3}$$

In this expression the slowly varying function T can be interpreted as the instantaneous temperature of the black hole. Indeed, from the conformal anomaly (see section 2.2), one finds an outgoing flux of quasithermal Hawking radiation

$$\langle{:}T_{uu}{:}\rangle \approx \frac{\pi c}{12} T(u)^2. \tag{4.3.4}$$

By integrating the first law one finds

$$S_{\mathrm{BH}}(v) \approx S_0 + \frac{2\pi\phi_r T(v)}{4G}, \tag{4.3.5}$$

which is linear in the temperature, as expected for a near-extremal black hole.

4.3.2 Backreaction

Now consider the effect of diary, modelled as an ingoing shockwave created a local quench in the matter sector, thrown in at time t_D. This leads to the profile for the left-moving component of the stress tensor (see section 2.5)

$$\langle{:}T_{vv}{:}\rangle = \frac{\Delta_D}{\pi} \frac{\varepsilon^2}{((v-t_D)^2+\varepsilon^2)^2} \to E_D\, \delta(v-t_D). \tag{4.3.6}$$

In this section we focus on the shockwave limit which arises in the limit where the regulator parameter $\varepsilon \to 0$ with the energy of the diary $E_D = \Delta_D/2\varepsilon$ held fixed. In section 4.4 we numerically study some aspects of finite width local quenches, that is, with ε small but finite. The diary also carries an entropy $S_D = \log d_D$, where d_D is the quantum dimension of the operator which created the shockwave [95–97].

In the presence of the shockwave the energy balance equation (2.3.7) reads

$$\frac{\mathrm{d}E}{\mathrm{d}t} = \langle{:}T_{vv}{:}\rangle - kE, \qquad E = -\frac{\phi_r}{8\pi G}\{F,t\}. \tag{4.3.7}$$

This is precisely the same equation for the function F that we had to solve in section 2.5 for the evaporating black hole solution. The only difference lies in the boundary conditions. Firstly, we have to account for the fact that the energy jumps by E_D as the diary is thrown in at $t = t_D$, and secondly, we have to match the solution at $t = t_D$,

by a Möbius transformation, onto the solution for the evaporating black hole (4.3.3) rather than the solution for the extremal black hole. As before, the exact solution can be expressed in terms of modified Bessel functions, however, there is simple WKB like solution which is valid until very late times

$$F(t) = \frac{\lambda + \exp\left[2\pi \int_{t_D}^{t} \mathrm{d}t' T(t')\right]}{1 + \lambda \exp\left[2\pi \int_{t_D}^{t} \mathrm{d}t' T(t')\right]}, \qquad \text{for } t \geq t_D, \tag{4.3.8}$$

where the temperature $T(t)$ is defined through the relation

$$E(t) = \frac{\pi \phi_r}{4G} T(t)^2. \tag{4.3.9}$$

Importantly, since E jumps discontinuously by E_D as the diary is thrown in, so does the temperature. This discontinuity in the temperature enters in the parameter

$$\lambda = \frac{T^+ - T^-}{T^+ + T-}, \qquad T^{\pm} = \lim_{t \to t_D^{\pm}} T(t), \tag{4.3.10}$$

of the Möbius transformation (4.3.8) which matches the two solutions across the shock-wave. The solution for the function $F(t)$ (4.3.8) in the presence of the diary suggests that we should introduce a new function

$$\widetilde{F}(t) = \exp\left(2\pi \int_{t_D}^{t} \mathrm{d}t'\, T(t')\right), \tag{4.3.11}$$

and a new set of Kruskal coordinates $\widetilde{U}$ and $\widetilde{V}$ defined by

$$\widetilde{U} = -1/\widetilde{F}(u), \qquad \widetilde{V} = \widetilde{F}(v). \tag{4.3.12}$$

Each of these new Kruskal coordinates are related to the old ones by a Möbius transformations

$$\widetilde{V} = \frac{V - \lambda}{1 - \lambda V}, \qquad \widetilde{U} = \frac{U + \lambda}{1 + \lambda U}. \tag{4.3.13}$$

This transformation is an isometry of AdS_2, so the metric has the same form (4.3.1) when written in terms of the new Kruskal coordinates.

The significance of these new coordinates is that they are adapted to the true event horizon of the black hole $\widetilde{U} = 0$, or $U = -\lambda$, and not the surface $U = 0$ which would have been the horizon if we hadn't thrown in the diary. It's tempting to say the diary causes the horizon to jump out as it is thrown into the black hole, however, this doesn't make sense since the event horizon is a teleological concept. In contrast, the apparent horizon does jump out after the diary is thrown in.

Using (4.3.5) we can compute the jump in the Bekenstein-Hawking entropy as the diary is thrown in

$$\Delta S_{\mathrm{BH}}(v) = S_{\mathrm{BH}}(t_D^+) - S_{\mathrm{BH}}(t_D^-) = S_{\mathrm{BH}}(t_D^-)\left(\frac{T^+}{T^-} - 1\right). \tag{4.3.14}$$

We can also rewrite the parameter λ (4.3.10) in terms of the Bekenstein-Hawking entropy

$$\lambda = \frac{\Delta S_{\mathrm{BH}}}{2S_{\mathrm{BH}}(t_D^-) + \Delta S_{\mathrm{BH}}}. \tag{4.3.15}$$

For simplicity, we will restrict to the case that the jump in the Bekenstein-Hawking entropy is much smaller than the original Bekenstein-Hawking entropy, that is, $\Delta S_{\mathrm{BH}} \ll S_{\mathrm{BH}}(t_D^-)$, which implies $\lambda \ll 1$, yet still large enough to have an interesting effect.

4.3.3 Islands

Consider the entropy of all the radiation collected up to time t, that is, we take the endpoint of R to lie on the boundary $u = v = t$. Assuming the black hole was formed by a heating up a zero-temperature black hole at $t = 0$, for early times, the entropy of R is given by [55]

$$S(R) = S_{\mathrm{rad}}(R) = \frac{\pi c}{6}\int_0^t T(t')\mathrm{d}t'. \tag{4.3.16}$$

Since the thermal entropy of the radiation grows monotonically with time and eventually it becomes favourable to have an island. To determine the location of the QES, we have to extremise the generalised entropy

$$S_{\mathrm{gen}}(I \cup R) = \frac{\phi_r}{4G}\left(\frac{F''(v)}{F'(v)} - \frac{1}{2}\frac{UF'(v)}{1+UV}\right) + \frac{c}{6}\log\frac{\sqrt{F'(v)F'(t)}}{U_{\partial R}^2}\frac{U - U_{\partial R}}{1+UV} + S_D(v). \tag{4.3.17}$$

The first term is the dilaton (4.3.1) evaluated at the QES, which we have parameterised in terms of Kruskal coordinates U, V as well as an ingoing coordinate v. The second term is the entropy of quantum fields on $I \cup R$ in the Unruh state. For simplicity, we've dropped the contribution to the entropy from the ingoing modes as these turn out not to play a role. We have also adopted the notation $U_{\partial R} = -1/F(t)$ for the location of the endpoint of R. The term

$$\frac{\sqrt{F'(v)F'(t)}}{U_{\partial R}^2}\frac{1}{1+UV}, \tag{4.3.18}$$

is the contribution from the Weyl anomaly, associated to the endpoints of I and R, which comes from rescaling the Minkowski metric in the reservoir and the AdS_2 metric to $-\mathrm{d}U\mathrm{d}v$. The final term in the generalised entropy (4.3.17) is the contribution from the entropy of the diary,

$$S_D(v) = S_D\Theta(v - t_D), \tag{4.3.19}$$

which only contributes if the diary lies in the island. Before the diary is thrown in, the generalised simplifies to

$$S_{\rm gen}(U,v) \approx S_{\rm BH}(v)(1-2UV) + \frac{c}{6}\log\left[\frac{\sqrt{F'(v)F'(t)}}{U_{\partial R}^2}(U-U_{\partial R})\right] + S_D(v). \quad (4.3.20)$$

Here we have assumed that the QES lies close to the horizon in the sense that $UV \ll 1$. This will be checked ex-post facto. Extremising the generalised entropy determines the location of the QES

$$U = -\frac{U_{\partial R}}{3}, \qquad V = -\frac{c}{16 S_{\rm BH}(v)}\frac{1}{U_{\partial R}}, \quad (4.3.21)$$

from which it is simple to see that the QES lies close to the horizon, as initially assumed.

We now turn to the entropy at the extremum. In the first instance, in addition to the leading order contributions that are of order $S_{\rm BH}$, we will keep track of logarithmic terms that are of order $c\log(S_{\rm BH}/c)$ and $c\log(S_{\rm BH}/\Delta S_{\rm BH})$. These contributions are associated to time scales of order the scrambling time. Note that in the logarithmic terms it doesn't matter whether the entropy and temperature are evaluated either at t_D or t, since the difference is beyond the order to which we are working. Henceforth, we will drop the arguments for these functions. Keeping only these terms, we find

$$\begin{aligned} &S_{\rm BH}(v) \approx S_{\rm BH}(t) + \frac{c}{24}\log\frac{S_{\rm BH}(t)}{c}, \\ &\frac{c}{6}\log\left[\frac{\sqrt{F'(v)F'(t)}}{U_{\partial R}^2}(U-U_{\partial R})\right] \approx -\frac{c}{12}\log\frac{S_{\rm BH}(t)}{c}, \end{aligned} \quad (4.3.22)$$

and so the generalised entropy is

$$S_{\rm gen}(I_1 \cup R) = S_{\rm BH}(t) - \frac{c}{24}\log\frac{S_{\rm BH}(t)}{c}. \quad (4.3.23)$$

Now suppose that t increases beyond the time t_D where the diary is thrown in. The solution (4.3.21) remains valid as long as $v < t_D$. For $v > t_D$ the solution for the function F jumps by a Möbius transformation, so we can no longer use the original solution (4.3.3).

For the following analysis, it will be important to note that for $u > t_D$, that is, $|\widetilde{U}_{\partial R}| < 1$, and small λ, we have

$$U_{\partial R} \approx \widetilde{U}_{\partial R} - \lambda. \quad (4.3.24)$$

This means that time scales where $t - t_D$ is much larger than the thermal scale, the logarithm of the RHS has a sharp crossover

$$\log(\lambda - \widetilde{U}_{\partial R}) \approx -\min\left\{\log\frac{S_{\rm BH}(t_D^-)}{\Delta S_{\rm BH}}, 2\pi\int_{t_0}^{t} T(t')dt'\right\}. \quad (4.3.25)$$

It then follows that the ingoing null coordinate v of the QES is approximately given by

$$v = \min\left\{ t - \frac{1}{2\pi T(t)} \log \frac{S_{\rm BH}(t)}{c}, t_D - \frac{1}{2\pi T^-} \log \frac{\Delta S_{\rm BH}}{c} \right\}. \tag{4.3.26}$$

This means that as time t increases, the ingoing null coordinate v of the QES eventually freezes to the past of the diary, so that $v < t_D$, with

$$v \to t_D - \frac{1}{2\pi T^-} \log \frac{\Delta S_{\rm BH}}{c}, \qquad U \to \frac{\Delta S_{\rm BH}}{6 S_{\rm BH}(t_D^-)}. \tag{4.3.27}$$

This is the same result as equation (92) in [10]. Notice that to have $v < t_D$ we need $\Delta S_{\rm BH} \geq c$, otherwise the QES never freezes to the past of the diary. This corresponds to the fact that for smaller diaries the effect of the backreaction is essentially negligible. Using this, we have

$$\begin{aligned} S_{\rm gen}(I_1 \cup R) = \max\Bigg\{ & S_{\rm BH}(t) - \frac{c}{24} \log \frac{S_{\rm BH}(t)}{c}, \\ & S_{\rm BH}(t_D^-) + \frac{c}{8} \log \frac{\Delta S_{\rm BH}}{c} - \frac{c}{6} \log \frac{S_{\rm BH}(t_D^-)}{c} + S_{\rm rad}(R_2) \Bigg\}. \end{aligned} \tag{4.3.28}$$

Here we have split the radiation R as $R_1 \cup R_2$ where R_1 is the radiation collected in the time interval $[-\infty, t_D]$ before the diary was thrown in and R_2 is the remaining radiation collected in the time interval $[t_D, t]$. In (4.3.28), the second term is eventually maximum and so the QES indeed freezes as described. If this was the only island then we would eventually run into a puzzle since $S_{\rm rad}(R_2)$ increases monotonically with time.

The resolution to this potential puzzle is that for later times $t > t_D$, after the diary has been thrown in, a new island I_2 with a QES that lies to the future of the diary appears (see figure 4.3). This solution for the new QES is simply (4.3.21) but written in the other set of Kruskal coordinates (4.3.13). Since the generalised entropy has the same functional form when written in terms of either set of Kruskal coordaintes, the only difference to (4.3.23) is that we need to include the entropy of the diary,

$$S_{\rm gen}(I_2 \cup R) = S_{\rm BH}(t) - \frac{c}{24} \log \frac{S_{\rm BH}(t)}{c} + S_D. \tag{4.3.29}$$

4.3.4 Information recovery

We are now in a position to understand when the information in the diary is recoverable in the radiation. By EWR, this happens when I_2 has minimal generalised entropy. For a diary thrown in past the Page time, this means

$$S_{\rm gen}(I_2 \cup R) \leq S_{\rm gen}(I_1 \cup R). \tag{4.3.30}$$

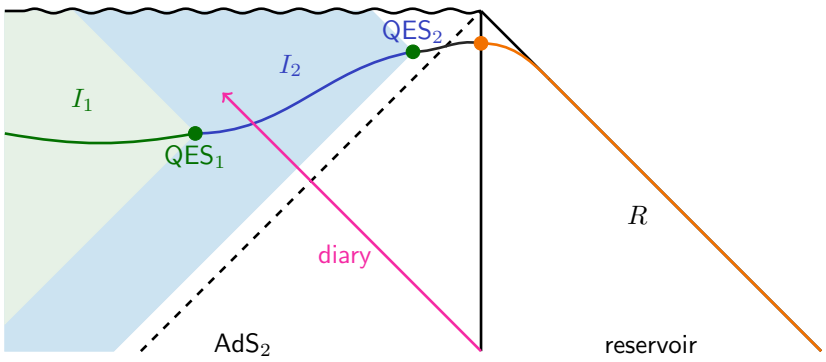


Figure 4.3: When a diary is thrown in from the boundary after the Page time, QES$_1$, which is responsible for the downward portion of the Page curve, gets stuck behind the diary. At later times, when QES$_2$ is the minimal QES, the diary lies in the entanglement wedge of the radiation.

Or, equivalently,

$$S_{\text{rad}}(R_2) \geq \frac{2}{3}\left(\Delta S_{\text{BH}} + S_D\right) + \frac{c}{12}\log\frac{S_{\text{BH}}}{\Delta S_{\text{BH}}}. \tag{4.3.31}$$

The delay between the time when the diary is thrown in and recovery time can be interpreted as the time it takes for the black hole to "process" the diary and return its information though Hawking radiation. The logarithmic terms correspond to a delay related to the scrambling time. Up to this logarithmic correction, this is precisely what we found from the islands in the stream recipe (4.2.11) with $\xi = 2$, which corresponds to the case of a trivial greybody factor or transparent boundary conditions.

4.4 Python's lunch

The analysis in the previous section of the generalised entropy in the shockwave limit shows that there are two competing minima of the generalised entropy S_{gen} which lie either side of $v = t_D$. In between the two minima there must be a maxima of S_{gen}. Since this maxima does not appear as a solution when we extremise the S_{gen} we infer that the maxima must lie at the point where S_{gen} is discontinuous, that is, at $v = t_D$. The discontinuity in $\partial_v S_{\text{gen}}$ arises from the fact that the energy jumps discontinuously as the diary is thrown in. More specifically, the discontinuity in $\partial_v S_{\text{gen}}$ arises from a discontinuity in $\partial_v \phi$ which can be seen from the solution (4.3.1). Whilst we demanded that the function F is continuous up to its second derivative when solving the Schwarzian equation, the third derivative of F is determined by the energy according to

$$E = -\frac{\phi_r}{8\pi G}\{F, t\}. \tag{4.4.1}$$

Since E has a discontinuity across $t = t_D$ this implies F must also have a discontinuity here. Away from the shockwave limit, that is, for finite ε, the generalised entropy is smooth and the maxima can be seen by extremising the generalised entropy. In this section we show how this works numerically.

Whilst maxima of the generalised entropy do not play any role as far as the entropy is concerned, they do play a role in another quantity: complexity. In the previous section we have seen how EWR allows us to understand when decoding task is theoretically possible, that is, when there exists a unitary operator on the radiation which recovers the state of the diary. However, it doesn't tell us anything about how complex this operator is. We can estimate the complexity using the Python's lunch conjecture [93]

$$\mathcal{C} \sim \exp\left(\frac{S_{\max} - S_{\min}}{2}\right). \tag{4.4.2}$$

Here $S_{\max}$ is the generalised entropy of the QES which is a maxima of S_{gen} whilst $S_{\min}$ is the generalised entropy of the neighbouring QES which is a minima of the generalised entropy.

In this section, we apply the Python's lunch conjecture to understand when it is possible to decode the diary alone. More specifically, while Harlow and Hayden [89] argued that it's exponentially hard to convert an old black hole and its radiation into a simple state, the question we ask is: assuming we're given this state, what is the complexity of decoding the diary?

4.4.1 Finding the maxima

We have already argued for the existence of a maxima of the generalised entropy in the shockwave limit. In particular, we found that it must lie along $v = t_D$. We now work away from the shockwave limit where left-moving component of the stress tensor is given by

$$\langle :T_{vv}: \rangle = \frac{\Delta_D}{\pi} \frac{\varepsilon^2}{((v - t_D)^2 + \varepsilon^2)^2}, \tag{4.4.3}$$

with ε small but finite. Although we do not have the analytic solution for the function F in this case, it turns out to be possible to obtain the exact solution for the outgoing null Kruskal coordinate U of the QES. We find

$$U = -\frac{k + kU_{\partial R}V + U_{\partial R}F'(v)}{kV + kU_{\partial R}V^2 - F'(v)}. \tag{4.4.4}$$

Using this we can then numerically study the "off shell"[4] generalised entropy as a function of the ingoing null coordinate v for various times t. This is shown in figure

[4]We use the term "off shell" since the generalised entropy is being evaluated on a surface which, for generic v, is not quantum extremal.

4.4 where we have also smoothed out the Heaviside function in $S_D(v)$ (4.3.19). The extremal points correspond to configurations which are extremal i.e. QESs. Initially there is only one minimal QES which lies to the past of the diary. However, for later times, a second QES appears which lies to the future of the diary. For sufficiently late times it is the latter QES which minimal generalised entropy and when this transition occurs the information in the diary is recoverable in the radiation. There must also be a QES which is maxima of the generalised entropy which lies in between, more specifically, it must lie in, or on the boundary of, the interval where the diary is localised. From numerics we find that it lies at the far edge of the region where the diary is localised.

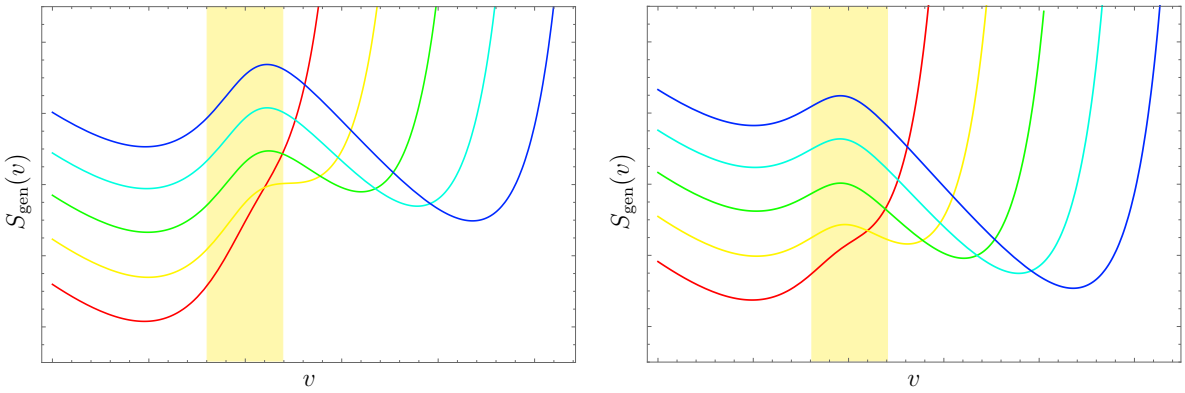


Figure 4.4: The off shell generalised entropy S_{gen} for R (left) and $R \cup \bar{D}$ (right), where $\bar{D}$ is a purification of D, for a finite width quench as a function of the ingoing null coordinate v of the QES for several values of t, with t increasing from red to blue. The energy density of the diary is localised in the region shaded in yellow. Here the parameters are: $k = 0.1, c = 10, \Delta_D = 10, h_D = 1, \varepsilon = 0.5, t_0 = 2/k$.

It's simple to understand why the maxima lies at the far edge of the interval where the diary is localised. The important terms in the generalised entropy which determine the ingoing null coordinate v of the QES, assuming it lies in the interval where the diary is localised, are

$$S_{\text{gen}} \supset S_{\text{BH}}(v) + S_D(v). \tag{4.4.5}$$

Here S_{BH} and S_D are monotonically increasing functions of v, due to the backreaction. Hence, for R the maximum will lie at the edge of the interval. The approximate coordinates of the QES at the recovery time are given by $v = t_D + \delta v$, with $\delta v = \mathcal{O}(\varepsilon)$, and U given by

$$U = U_{\partial R} + \frac{c}{12 S_{\text{BH}}}, \tag{4.4.6}$$

which follows from (4.4.4). Hence, at the maximum, the generalised entropy is

$$S_{\max} = S_{\text{BH}}(t_D + \delta v) + S(D) + S_{\text{rad}}(R_2) = S_{\text{BH}}(t_D) + \Delta S_{\text{BH}} + S(D) + S_{\text{rad}}(R_2). \tag{4.4.7}$$

Using the Python's lunch conjecture (4.4.2) this gives

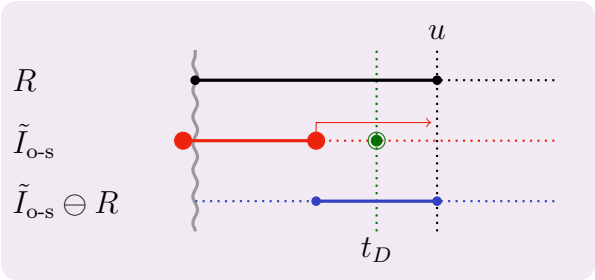
$$\mathcal{C} \sim \exp\left(\frac{\Delta S_{\text{BH}} + S_D}{2}\right). \tag{4.4.8}$$

In the reversible limit, that is, when the GSL $\Delta S_{\text{BH}} \geq S_D$ is saturated, the decoding complexity is simply determined by the dimension of the code space, that is, e^{S_D}. In this limit the result is in agreement with arguments based on random unitary models [98]. Notice that the irreversibility of the process of the black hole absorbing the diary increases the complexity of decoding.

4.4.2 Islands in the stream and Python's lunch

We now discuss a simplified way of deriving the complexity of decoding using the island in the stream recipe, which was discussed in section 4.2. The idea is to consider an off shell *sweep* of the generalised entropy $S_{\text{gen}}(I_{\text{o-s}} \cup R)$, that is, where $\partial I_{\text{o-s}}$ are not necessarily QESs [93]. In the results of the islands in the stream recipe, there is a natural way to define the off shell generalised entropy: simply take the islands in the stream formula but do not insist that the endpoints of $\partial \tilde{I}$ correspond to points in ∂R. Varying the one of the endpoints $\partial \tilde{I}$ gives rise to a sweep.

For a diary is thrown into a black hole let's consider the sweep of the generalised entropy with an off-shell island $\tilde{I}_{\text{o-s}} = [0, u_{o-s}]$:



The off shell island has one endpoint located at the point just before the black hole was formed by an infalling shockwave. The sweep is shown in figure 4.5. The generalised entropy jumps by $S_{\text{BH}}(0)$ as $\tilde{u}_{\text{o-s}}$ crosses 0. It then decreases until the diary is thrown in at $\tilde{u}_{\text{o-s}} = t_D$. Here, the generalised entropy has another minimum, which corresponds to the QES which gets stuck to the past of the diary. For late enough times, a QES which lies to the future of the diary has minimal generalised entropy. Between these two minima, there is a maxima located at the point just after the diary was thrown in.

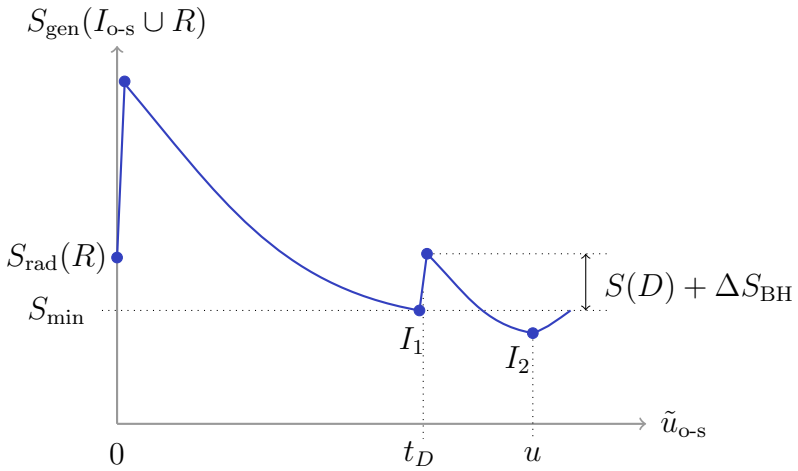


Figure 4.5: An off shell sweep of the generalised entropy with $\tilde{I}_{\text{o-s}} = [0, u_{\text{o-s}}]$. The blue dots correspond to the extrema of the generalised entropy. The sweep reveals a local maxima which determines the complexity of decoding the diary.

4.5 Petz map

In this section we study the problem of information recovery in a toy model, which is closely related to the random isometry model introduced in section 3.7.4, using the Petz map. For clarity we begin by working in a simpler, coarse-grained version of the model.

4.5.1 A simple model

In this section we consider a simple toy model of a black hole in a holographic setup, defined by a linear map

$$V = d_r \, \langle \text{MAX}|_{r'r} \, U \, |\psi_0\rangle_f \, , \tag{4.5.1}$$

from the Hilbert space $\mathcal{H}_\ell \otimes \mathcal{H}_r$ of left-moving and right-moving interior modes to the Hilbert space $\mathcal{H}_B$ of black hole microstates. Here $|\psi_0\rangle_f$ is a fixed state in some ancillary system $\mathcal{H}_f$, $|\text{MAX}\rangle_{r'r}$ denotes the maximally entangled state on $\mathcal{H}_{r'} \otimes \mathcal{H}_r$, where $\mathcal{H}_{r'}$ is an auxiliary Hilbert space isomorphic to $\mathcal{H}_r$, and $U : \mathcal{H}_\ell \otimes \mathcal{H}_f \to \mathcal{H}_B \otimes \mathcal{H}_{r'}$ is a Haar random unitary. The map V is shown in figure 4.6 as the part enclosed by blue dashed lines. This is a coarse-grained version of the dynamical model introduced in section 7 of [99]. Some aspects of this model have been studied in [100].

To model an evaporating black hole we entangle the right-moving interior modes with an auxiliary reference system $\mathcal{H}_R$, which models the reservoir into which the black hole evaporates. The overall state of the system is

$$|\Psi\rangle = (V \otimes \text{id}_R) \, |\psi\rangle \, , \tag{4.5.2}$$

where $|\psi\rangle$ is a pure bulk state on $\mathcal{H}_\ell \otimes \mathcal{H}_r \otimes \mathcal{H}_R$. To understand the motivation for this model, consider taking the bulk state $|\psi\rangle$ to be maximally entangled on $\mathcal{H}_r \otimes \mathcal{H}_R$, as shown in figure 4.6. The maximally entangled state is the analogue of the Unruh state (2.2.6) in this toy model. By bending around the green line in figure 4.6 we see that this precisely coincides with Page's model (figure 3.6).

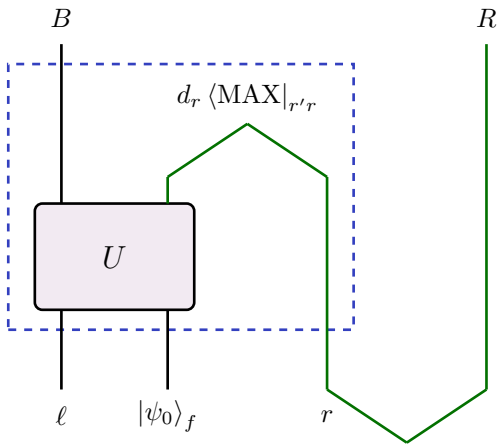


Figure 4.6: A coarse-grained version of the model in [99].

4.5.2 Reconstruction in the simple model

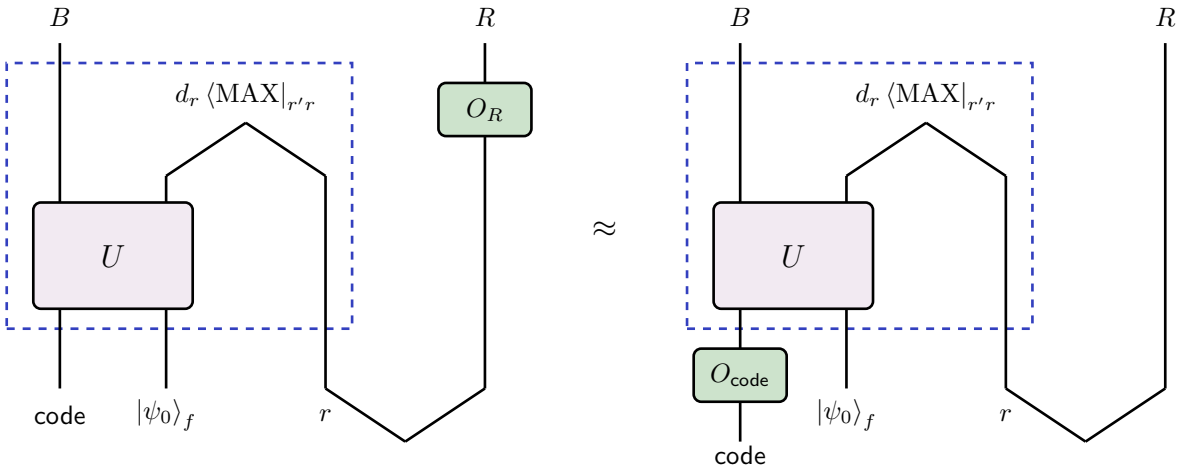
In this section we consider the problem of information recovery for the left-moving modes which can naturally be identified as the diary system. To this end, consider the code subspace $\mathcal{H}_{\text{code}} = \mathcal{H}_\ell \otimes \mathbb{C}\,|\mathrm{MAX}\rangle_{rR}$. The image of $\mathcal{H}_{\text{code}}$ under $V \otimes \mathrm{id}_R$ is spanned by the states

$$|\Psi_a\rangle = \frac{1}{\sqrt{d_R}} \sum_{i=1}^{d_R} |\psi_{ai}\rangle_B \otimes |i\rangle_R\,, \quad a = 1, 2, \ldots, d_{\text{code}}, \tag{4.5.3}$$

where $|\psi_{ai}\rangle_B = V\,|a\rangle_\ell \otimes |i\rangle_r$ is a black hole microstate corresponding to the left-moving and right-moving interior modes being in the product state $|a\rangle_\ell \otimes |i\rangle_r$. Notice that the encoding $\mathcal{H}_{\text{code}} \to \mathcal{H}_B \otimes \mathcal{H}_R$ happens to be an isometry. For generic code spaces $\mathcal{H}_{\text{code}} \subseteq \mathcal{H}_\ell \otimes \mathcal{H}_r \otimes \mathcal{H}_R$ this won't be the case. Let us denote the encoding $\iota : \mathcal{H}_{\text{code}} \to \mathcal{H}_B \otimes \mathcal{H}_R$ and define

$$O_{BR} \coloneqq \iota\, O_{\text{code}}\, \iota^\dagger = \sum_{a,b=1}^{d_{\text{code}}} O_{ab} |\Psi_a\rangle\langle\Psi_b|, \tag{4.5.4}$$

where O_{ab} are the matrix elements of O_{code}. The operator O_{BR} provides an *exact global reconstruction* of O_{code}. It is exact because the encoding is an isometry and global because it acts jointly on B and R. The question we are interested in is: does there



exist a reconstruction O_{code} which only acts on R? That is, does there exist an operator O_R on R such that:

By a slight abuse of notation, in this diagram, we are identifying $\mathcal{H}_{\mathrm{code}}$ with $\mathcal{H}_\ell$, since they are isomorphic. This problem is nothing but the problem of information recovery stated in the Heisenberg picture, where we talk about reconstructing operators, rather than the Schrödinger picture, where we talk about recovering the state of the diary.

There is a theorem by Dénes Petz [92] in the theory of recovery channels which says that when reconstruction is possible we can take the reconstruction to be given by[5]

$$O_R = \frac{1}{d_{\mathrm{code}}} \sigma_R^{-1/2} \operatorname{Tr}_B(O_{BR})\, \sigma_R^{-1/2}, \tag{4.5.5}$$

which is known as the Petz map. Here $\sigma_R = \operatorname{Tr}_B(\iota\iota^\dagger)/d_{\mathrm{code}}$. Strictly speaking, this theorem applies in the exact setting, that is, where exact quantum error correction is possible. In the approximate setting, the Petz map does not work perfectly, however, it has an average error which is almost as small as the average error of the optimal reconstruction [101].

Actually, due to the simplicity of the model considered here, the simpler *Petz lite* [11],

$$O_R = c_0 \operatorname{Tr}_B O_{BR}, \tag{4.5.6}$$

is expected to work. Here c_0 is a constant chosen so that the map preserves the identity element. More specifically, the Petz lite map is supposed to work if the model has fixed

[5] To see why this works, suppose that ι forms an exact subsystem code for R. Harlow's structure theorem (theorem 4.1 in [31]) implies that there exists an isomorphism $W_R : \mathcal{H}_R \to \mathcal{H}_{R_1} \otimes \mathcal{H}_{R_2}$, with $\mathcal{H}_{\mathrm{code}}$ isomorphic to a subspace of $\mathcal{H}_{R_2}$, such that $(\mathrm{id}_B \otimes W_R)\iota\,|\psi\rangle_{\mathrm{code}} = |\chi\rangle_{BR_1} \otimes |\psi\rangle_{R_2}$, for some fixed state $|\chi\rangle_{BR_1}$. A short calculation reveals that (4.5.5) provides a reconstruction.

area in the sense of [102, 103]. It's now straightforward to check when O_R provides a good reconstruction by calculating the matrix elements

$$\langle\Psi_a|\, O_R\, |\Psi_b\rangle = \frac{c_0}{d_R^2} \sum_{c,d=1}^{d_{\text{code}}} O_{cd} \sum_{i,j=1}^{d_R} \langle\psi_{ai}|\psi_{bj}\rangle \langle\psi_{dj}|\psi_{ci}\rangle\,. \tag{4.5.7}$$

The average of the matrix elements can be computed using the formula

$$\begin{aligned}\mathbb{E}\left[U_{\mu_1\nu_1}U_{\mu_2\nu_2}U^\dagger_{\nu_1'\mu_1'}U^\dagger_{\nu_2'\mu_2'}\right] = \frac{1}{N^2-1}\Big[&\delta_{\mu_1\mu_1'}\delta_{\mu_2\mu_2'}\delta_{\nu_1\nu_1'}\delta_{\nu_2\nu_2'} + \delta_{\mu_1\mu_2'}\delta_{\mu_2\mu_1'}\delta_{\nu_1\nu_2'}\delta_{\nu_2\nu_1'} \\ &- \frac{1}{N}\left(\delta_{\mu_1\mu_1'}\delta_{\mu_2\mu_2'}\delta_{\nu_1\nu_2'}\delta_{\nu_2\nu_1'} + \delta_{\mu_1\mu_2'}\delta_{\mu_2\mu_1'}\delta_{\nu_1\nu_1'}\delta_{\nu_2\nu_2'}\right)\Big],\end{aligned} \tag{4.5.8}$$

where the average is taken over the Haar measure on the unitary group $U(N)$. We find

$$\mathbb{E}\,\langle\Psi_a|\, O_R\, |\Psi_b\rangle = c_1 O_{ab} + c_2 \mathrm{Tr}\, O_{\text{code}}\delta_{ab}, \tag{4.5.9}$$

with

$$c_1 = \frac{1}{1+\lambda}, \qquad c_2 = \frac{1}{d_{\text{code}}}\frac{\lambda}{1+\lambda}, \qquad \lambda \approx \frac{d_B d_{\text{code}}}{d_R}. \tag{4.5.10}$$

That c_1 and c_2 are determined by a single parameter λ follows from the constraint $c_1 + d_{\text{code}}c_2 = 1$, which ensures that the map preserves the identity element. We see that the reconstruction works on average when $d_R > d_B d_{\text{code}}$, rather than after the Page time. Hence, even long after the Page time, reconstruction is only possible for sufficiently small code spaces. This is essentially the decoding criterion for the case that the diary was thrown into a young black hole, that is, a black hole before the Page time. The delay related to the dimension of the code space is in line with expectations [47, 82, 94] and is related to the fact that entanglement wedge reconstruction is state dependent. For simplicity, we dropped corrections to λ which are suppressed when the dimensions of the subsystems are large. The exact expression

$$\lambda = \frac{d_B\left(1-\frac{1}{d_B^2}\right)d_{\text{code}}}{d_R\left(1-\frac{1}{d_R^2}\right)} \tag{4.5.11}$$

is only necessary for very early or very late times when $d_R \to 1$ or $d_B \to 1$.

One way to generalise this calculation is to include some of the right-moving interior modes in the code subspace. To this end, we split $\mathcal{H}_r = \mathcal{H}_{r_1} \otimes \mathcal{H}_{r_2}$ and consider the code subspace $\mathcal{H}_{\text{code}} = \mathcal{H}_\ell \otimes \mathcal{H}_{r_1} \otimes \mathbb{C}\,|\mathrm{MAX}\rangle_{r_2 R}$. Whilst the encoding is no longer an isometry, on average we find

$$\mathbb{E}\,\langle\Psi_A|\, O_R\, |\Psi_B\rangle = c_1' O_{AB} + c_2'\left(\mathrm{Tr}_{r_1} O_{\text{code}}\right)_{ab}\delta_{i_1 j_1} + c_3'\left(\mathrm{Tr}_\ell\, O_{\text{code}}\right)_{i_1 j_1}\delta_{ab} + c_4' \mathrm{Tr}\, O_{\text{code}}\delta_{AB}, \tag{4.5.12}$$

with

$$c_1' \approx \frac{1}{1+\lambda'}, \quad c_{2,3}' = O(d_B^{-\#}), \quad c_4' \approx \frac{1}{d_{\text{code}}}\frac{\lambda'}{1+\lambda'}, \quad \lambda' = \frac{d_B d_{\text{code}}}{d_{r_2}}. \tag{4.5.13}$$

In (4.5.12) we introduced the composite indices $A = (a, i_1)$ and $B = (b, j_1)$. For simplicity, we dropped corrections which are suppressed when the dimensions of the subsystems are large and have not written the formula for c_2' and c_3' as they are exponentially suppressed in $\log d_B$. As a simple consistency check, for $d_{r_1} \to 1$ we recover (4.5.9) with $c_1 = c_1' + c_2'$ and $c_2 = c_3' + c_4'$.

4.5.3 A dynamical model

In this section we will consider the dynamical model defined in section 7 of [99]. The map $V_k : \mathcal{H}_{\ell_k} \otimes \mathcal{H}_{r_k} \to \mathcal{H}_{B_k}$ is defined recursively, starting with

$$V_0 = U_0 \left|\psi_0\right\rangle_f, \tag{4.5.14}$$

and then at each time step acting with a unitary U_k followed by post-selecting on the maximally entangled state

$$V_k = d_{r_k} \left\langle \text{MAX}\right|_{r_k' r_k} U_k U_{k-1} \dots U_0 \left|\psi_0\right\rangle_f. \tag{4.5.15}$$

As before, $\left|\psi_0\right\rangle_f$ is a fixed state in some ancillary system $\mathcal{H}_f$, $\mathcal{H}_{r_k'}$ is an auxiliary Hilbert space isomorphic to $\mathcal{H}_{r_k}$ and $U_k : \mathcal{H}_{\ell_{\{k\}}} \otimes \mathcal{H}_{B_{k-1}} \to \mathcal{H}_{B_k} \otimes \mathcal{H}_{r'_{\{k\}}}$ is a Haar random unitary, for each k. When acting on input bulk states which are maximally entangled on $\mathcal{H}_{r_{\{k\}}} \otimes \mathcal{H}_{R_{\{k\}}}$, up to notation, this model precisely corresponds to the random isometry model introduced in section 3.7.4. To clarify the notation used here, we denote by $\mathcal{H}_{R_{\{k\}}}$ the Hilbert space of the radiation emitted at the k^{th} time step and use the shorthand $\mathcal{H}_{R_k} = \mathcal{H}_{R_{\{k,\dots,1\}}} = \mathcal{H}_{R_{\{k\}}} \otimes \dots \otimes \mathcal{H}_{R_{\{1\}}}$ for the Hilbert space of the radiation emitted in the first k time steps. We adopt a similar notation for the left-moving and right-moving interior modes.

4.5.4 Recursion relations

In this section we describe a method to recursively compute averages involving products of V_k. In the following section we apply this method compute the matrix elements of the Petz Lite map. Let $\pi \in S_n$ and consider the average

$$\mathsf{X}_k^\pi := \mathbb{E}\left[(V_k^\dagger)^{\otimes n} \pi V_k^{\otimes n}\right]. \tag{4.5.16}$$

A closely related quantity was considered in section 3.7.4. It follows from the recursive definition of V_k that X_k^π satisfies the recursion relation

$$\mathsf{X}_k^{\pi_1} = \sum_{\pi_2 \in S_n} M_k(\pi_1, \pi_2) \mathsf{X}_{k-1}^{\pi_2}, \tag{4.5.17}$$

where $M_k(\pi_1, \pi_2)$ may be written explicitly in terms of the Weingarten function for the unitary group. It's convenient to define the vector

$$|\mathsf{X}_k\rangle = \sum_{\pi \in S_n} \mathsf{X}_k^\pi \otimes |\pi\rangle\,, \tag{4.5.18}$$

where, to keep track of the permutation, we tensor in $|\pi\rangle$ which belongs to a Hilbert space with an orthonormal basis consisting of vectors $|\pi\rangle$ with $\pi \in S_n$. It's then natural to view the operator $M_k(\pi_1, \pi_2)$ as the matrix elements of an operator

$$M_k = \sum_{\pi_1, \pi_2 \in S_n} M_k(\pi_1, \pi_2) |\pi_1\rangle\langle\pi_2|, \tag{4.5.19}$$

which acts on a bigger Hilbert space. Equipped with this notation, the recursion relation (4.5.17) reads

$$|\mathsf{X}_k\rangle = M_k\,|\mathsf{X}_{k-1}\rangle\,, \tag{4.5.20}$$

which implies $|\mathsf{X}_k\rangle = M_k M_{k-1} \dots M_1\,|\mathsf{X}_0\rangle$. In the following section we will be interested in the case that $n = 2$ since the Petz lite map involves the product of two matrix elements, that is, it involves two V's and two $V^\dagger$'s. For this case, the matrix elements of M_k are:

$$\begin{aligned}
M_k(\mathrm{id}, \mathrm{id}) &= \frac{d_{B_k} d_{R_{\{k\}}}}{d_{B_k}^2 d_{R_{\{k\}}}^2 - 1} \left(d_{B_k} d_{R_{\{k\}}} \mathrm{id} - \frac{1}{d_{B_k}} \mathrm{id} \otimes \tau \right), \\
M_k(\mathrm{id}, \tau) &= \frac{d_{B_k} d_{R_{\{k\}}}}{d_{B_k}^2 d_{R_{\{k\}}}^2 - 1} \left(d_{R_{\{k\}}} \tau - \tau \otimes \mathrm{id} \right), \\
M_k(\tau, \mathrm{id}) &= \frac{d_{B_k} d_{R_{\{k\}}}}{d_{B_k}^2 d_{R_{\{k\}}}^2 - 1} \left(d_{R_{\{k\}}} \mathrm{id} - \mathrm{id} \otimes \tau \right), \\
M_k(\tau, \tau) &= \frac{d_{B_k} d_{R_{\{k\}}}}{d_{B_k}^2 d_{R_{\{k\}}}^2 - 1} \left(d_{B_k} d_{R_{\{k\}}} \tau - \frac{1}{d_{B_k}} \tau \otimes \mathrm{id} \right).
\end{aligned} \tag{4.5.21}$$

In this expression, the first factor acts on $\mathcal{H}_{\ell_{\{k\}}} \otimes \mathcal{H}_{\ell_{\{k\}}}$ and the second factor acts on $\mathcal{H}_{r_{\{k\}}} \otimes \mathcal{H}_{r_{\{k\}}}$, so, for example $\mathrm{id} \otimes \tau$ is the identity on $\mathcal{H}_{\ell_{\{k\}}} \otimes \mathcal{H}_{\ell_{\{k\}}}$ and the cyclic permutation (or SWAP) operator on $\mathcal{H}_{r_{\{k\}}} \otimes \mathcal{H}_{r_{\{k\}}}$.

4.5.5 Reconstruction in the dynamical model

In this section we consider the problem of information recovery for diary, which is identified with the left-moving interior modes $\ell_{\hat{k}}$, thrown in at the $\hat{k}^{\text{th}}$ time step. To this end, consider the subspace spanned by the states

$$|\Psi_a\rangle = \frac{1}{\sqrt{d_{R_K}}} \sum_{i_1, \dots, i_K} |\psi_{a i_1 \dots i_K}\rangle_{B_K} \otimes |i_1 \dots i_K\rangle_{R_K}\,, \quad a = 1, 2, \dots, d_{\text{code}}, \tag{4.5.22}$$

which describes a code subspace of a subset of the left-moving interior modes at the $\hat{k}^{\text{th}}$ time step labelled by the index a, with the rest fixed. In particular, we take the state on the rest of the left-moving interior modes to be some fixed product state, whose role is only make the process of evaporation thermodynamically irreversible. Notice that the encoding happens to be an isometry, as was the case for the first example considered in section 4.5.2.

We can check when the Petz Lite map (4.5.6) works on average by computing

$$\mathbb{E}\,\langle\Psi_a|\,O_R\,|\Psi_b\rangle = \frac{c_0}{d_{R_K}^2}\sum_{c,d=1}^{d_{\text{code}}} O_{cd}\sum_{\mathsf{i},\mathsf{j}}\mathbb{E}\,\langle\psi_{a\mathsf{i}}|\psi_{b\mathsf{j}}\rangle\,\langle\psi_{d\mathsf{j}}|\psi_{c\mathsf{i}}\rangle\,, \tag{4.5.23}$$

where we introduced the composite indices $\mathsf{i} = (i_1, \ldots, i_K)$ and $\mathsf{j} = (j_1, \ldots, j_K)$. Using (4.5.16), the average of the matrix elements in this expression can be written as

$$\mathbb{E}\,\langle\psi_{a\mathsf{i}}|\psi_{b\mathsf{j}}\rangle\,\langle\psi_{d\mathsf{j}}|\psi_{c\mathsf{i}}\rangle = \langle a\mathsf{i}|\otimes\langle d\mathsf{j}|\,\mathsf{X}_K^{\text{id}}\,|b\mathsf{j}\rangle\otimes|c\mathsf{i}\rangle\,. \tag{4.5.24}$$

Because of the sum over i and j in the Petz lite map, we are actually interested in computing

$$\sum_{\mathsf{i},\mathsf{j}}\mathbb{E}\,\langle\psi_{a\mathsf{i}}|\psi_{b\mathsf{j}}\rangle\,\langle\psi_{d\mathsf{j}}|\psi_{c\mathsf{i}}\rangle = \langle a|\otimes\langle d|\,\mathrm{Tr}_{r_K}[\mathsf{X}_K^{\text{id}}]\,|b\rangle\otimes|c\rangle\,. \tag{4.5.25}$$

Using the recursion (4.5.20) we can write $\mathsf{X}_K^{\text{id}} = \langle\text{id}|\mathsf{X}_K\rangle$ as

$$\mathsf{X}_K^{\text{id}} = \langle\text{id}|\,M_K\ldots M_{\hat{k}}\ldots M_1\,|\mathsf{X}_0\rangle\,. \tag{4.5.26}$$

It's then useful to think of the partial trace Tr_{r_K} over the right-moving interior modes in the first K time steps in (4.5.25) as a product of partial traces at each time step, that is,

$$\mathrm{Tr}_{r_K} = \bigotimes_k \mathrm{Tr}_{r_{\{k\}}}. \tag{4.5.27}$$

This allows use to rewrite (4.5.25) as

$$\frac{1}{d_{R_K}^2}\sum_{\mathsf{i},\mathsf{j}}\mathbb{E}\,\langle\psi_{a\mathsf{i}}|\psi_{b\mathsf{j}}\rangle\,\langle\psi_{d\mathsf{j}}|\psi_{c\mathsf{i}}\rangle = \langle\text{id}|\,\mathsf{M}_K\ldots\langle a|\otimes\langle d|\,\mathsf{M}_{\hat{k}}\,|b\rangle\otimes|c\rangle\ldots\mathsf{M}_1\,|\mathsf{X}_0\rangle\,, \tag{4.5.28}$$

where M_k is proportional to the partial trace of M_k over $r_{\{k\}}$, that is,

$$\mathsf{M}_k = \frac{1}{d_{r_{\{k\}}}^2}\mathrm{Tr}_{r_{\{k\}}}[M_k]. \tag{4.5.29}$$

As long as the diary was not thrown in at the 0^{th} time step, we have $|\mathsf{X}_0\rangle = \sum_{\pi\in S_2}|\pi\rangle$, so (4.5.28) can be written as

$$\frac{1}{d_{R_K}^2}\sum_{\mathsf{i},\mathsf{j}}\mathbb{E}\,\langle\psi_{a\mathsf{i}}|\psi_{b\mathsf{j}}\rangle\,\langle\psi_{d\mathsf{j}}|\psi_{c\mathsf{i}}\rangle = \sum_{\pi\in S_2}\langle\text{id}|\,\mathsf{M}_K\ldots\langle a|\otimes\langle d|\,\mathsf{M}_{\hat{k}}\,|b\rangle\otimes|c\rangle\ldots\mathsf{M}_1\,|\pi\rangle\,. \tag{4.5.30}$$

To compute this we just need to do the matrix multiplication of all the M_k's and then sandwich them between the bra $\langle \mathrm{id}|$ and kets $|\mathrm{id}\rangle$ and $|\tau\rangle$. The matrix elements of M_k can be straightforwardly be obtained from (4.5.21). For $k \neq \hat{k}$ we find:

$$\begin{aligned}
\mathsf{M}_{\mathsf{k}}(\mathrm{id},\mathrm{id}) &= \frac{d_{B_k} d_{R_{\{k\}}}}{d_{B_k}^2 d_{R_{\{k\}}}^2 - 1} d_{B_k} \left(1 - \frac{1}{d_{B_k}^2}\right), \\
\mathsf{M}_{\mathsf{k}}(\mathrm{id},\tau) &= \frac{d_{B_k} d_{R_{\{k\}}}}{d_{B_k}^2 d_{R_{\{k\}}}^2 - 1} d_{R_{\{k\}}} \left(1 - \frac{1}{d_{R_{\{k\}}}^2}\right), \\
\mathsf{M}_{\mathsf{k}}(\tau,\mathrm{id}) &= 0, \\
\mathsf{M}_{\mathsf{k}}(\tau,\tau) &= \frac{d_{B_k} d_{R_{\{k\}}}}{d_{B_k}^2 d_{R_{\{k\}}}^2 - 1} d_{B_k} d_{R_{\{k\}}} \left(1 - \frac{1}{d_{B_k}^2 d_{R_{\{k\}}}^2}\right).
\end{aligned} \tag{4.5.31}$$

For $k = \hat{k}$, we instead have

$$\begin{aligned}
\langle a| \otimes \langle d|\, \mathsf{M}_{\hat{k}}(\mathrm{id},\mathrm{id})\, |b\rangle \otimes |c\rangle &= \frac{d_{B_{\hat{k}}} d_{R_{\{\hat{k}\}}}}{d_{B_{\hat{k}}}^2 d_{R_{\{\hat{k}\}}}^2 - 1} d_{B_{\hat{k}}} \left(1 - \frac{1}{d_{B_{\hat{k}}}^2}\right) \delta_{ab}\delta_{cd}, \\
\langle a| \otimes \langle d|\, \mathsf{M}_{\hat{k}}(\mathrm{id},\tau)\, |b\rangle \otimes |c\rangle &= \frac{d_{B_{\hat{k}}} d_{R_{\{\hat{k}\}}}}{d_{B_{\hat{k}}}^2 d_{R_{\{\hat{k}\}}}^2 - 1} d_{R_{\{\hat{k}\}}} \left(1 - \frac{1}{d_{R_{\{\hat{k}\}}}^2}\right) \delta_{ad}\delta_{bc}, \\
\langle a| \otimes \langle d|\, \mathsf{M}_{\hat{k}}(\tau,\mathrm{id})\, |b\rangle \otimes |c\rangle &= 0, \\
\langle a| \otimes \langle d|\, \mathsf{M}_{\hat{k}}(\tau,\tau)\, |b\rangle \otimes |c\rangle &= \frac{d_{B_{\hat{k}}} d_{R_{\{\hat{k}\}}}}{d_{B_{\hat{k}}}^2 d_{R_{\{\hat{k}\}}}^2 - 1} d_{B_{\hat{k}}} d_{R_{\{\hat{k}\}}} \left(1 - \frac{1}{d_{B_{\hat{k}}}^2 d_{R_{\{\hat{k}\}}}^2}\right) \delta_{ad}\delta_{bc}.
\end{aligned} \tag{4.5.32}$$

We see that for each k, M_k is an upper triangular matrix. This makes (4.5.30) particularly straightforward to evaluate. For the purpose of this calculation, it is convenient to write

$$\mathsf{M}_k = \begin{pmatrix} \mathsf{A}_k & \mathsf{B}_k \\ 0 & \mathsf{C}_k \end{pmatrix} \quad k \neq \hat{k}, \qquad \langle a| \otimes \langle d|\, \mathsf{M}_{\hat{k}}\, |b\rangle \otimes |c\rangle = \begin{pmatrix} \mathsf{A}_{\hat{k}}\delta_{ab}\delta_{cd} & \mathsf{B}_{\hat{k}}\delta_{ad}\delta_{bc} \\ 0 & \mathsf{C}_{\hat{k}}\delta_{ad}\delta_{bc} \end{pmatrix}. \tag{4.5.33}$$

The two terms appearing in the sum on the RHS of (4.5.30) are given by

$$\langle \mathrm{id}|\, \mathsf{M}_K \ldots \mathsf{M}_1\, |\mathrm{id}\rangle = \prod_{k=1}^{K} A_k\, \delta_{ab}\delta_{cd}, \tag{4.5.34}$$

and

$$\langle \mathrm{id}|\, \mathsf{M}_K \ldots \mathsf{M}_1\, |\tau\rangle = \sum_{k<\hat{k}} \prod_{l=k+1}^{K} A_l\, B_k \prod_{m=k-1}^{K} C_m\, \delta_{ab}\delta_{cd} + \sum_{k\geq\hat{k}} \prod_{l=k+1}^{K} A_l\, B_k \prod_{m=k-1}^{K} C_m\, \delta_{ad}\delta_{bc}. \tag{4.5.35}$$

This means that, as in the simple model, we can write

$$\mathbb{E}\,\langle\Psi_a|\,O_R\,|\Psi_b\rangle = c_1 O_{ab} + c_2 \mathrm{Tr}\, O_{\text{code}}\delta_{ab}, \tag{4.5.36}$$

with c_1 and c_2 parameterised as

$$c_1 = \frac{1}{1+\lambda}\,, \qquad c_2 = \frac{1}{d_{\text{code}}}\frac{\lambda}{1+\lambda}. \tag{4.5.37}$$

A little algebra reveals that the parameter λ may be written exactly as

$$\frac{\lambda}{d_{\text{code}}} = \frac{1+\sum_{k<\hat{k}} q_k}{\sum_{k\geq\hat{k}} q_k}, \qquad q_k = \zeta_k \frac{d_{R_k}}{d_{B_k}}, \tag{4.5.38}$$

with

$$\zeta_k = \frac{1-d_{R_{\{k\}}}^{-2}}{1-d_{B_k}^{-2}} \prod_{l=1}^{k-1} \frac{1-d_{B_l}^{-2}d_{R_{\{l\}}}^{-2}}{1-d_{B_l}^{-2}}, \tag{4.5.39}$$

which, up to the factor of $1-d_{R_{\{k\}}}^{-2}$, is equal to one plus exponentially small corrections in $\log d_{B_k}$. In the limit where the dimension of the black hole system is large, we can approximate the sums in (4.5.38) as

$$\sum_{k<\hat{k}} q_k = a_- \frac{d_{R_{\hat{k}-1}}}{d_{B_{\hat{k}-1}}}, \qquad \sum_{k\geq\hat{k}} q_k = a_+ \frac{d_{R_K}}{d_{B_K}}, \tag{4.5.40}$$

where the constants $a_\pm$ are give by

$$a_- \approx 1 + \sum_{k<\hat{k}-1} \frac{d_{B_{\hat{k}-1}}}{d_{B_k}} \frac{d_{R_k}}{d_{R_{\hat{k}-1}}}, \qquad a_+ \approx 1 + \sum_{k\geq\hat{k}+1} \frac{d_{B_K}}{d_{B_k}} \frac{d_{R_k}}{d_{R_K}}. \tag{4.5.41}$$

To keep things simple we have ignored the term $1-d_{R_{\{k\}}}^{-2}$ in ζ_k, although it is simple to put this back. Up to this unimportant point, $\approx$ means up to exponentially small corrections in $\min\{\log d_{B_{k-1}}, \log d_{B_K}\}$ which corresponds to $\min\{S_{\text{BH}}^{\text{before diary}}, S_{\text{BH}}\}$. We can place a simple upper bound on the parameters $a_\pm$

$$a_- \lesssim 1 + (\hat{k}-2)2^{\Delta S_{B_{\hat{k}-1}} - \Delta S_{R_{\hat{k}-1}}}, \qquad a_+ \lesssim 1 + (K-\hat{k}+1)2^{\Delta S_{B_K} - \Delta S_{R_K}}, \tag{4.5.42}$$

where we have defined

$$\Delta S_{B_k} = \log d_{B_k} - \log d_{B_{k-1}}, \qquad \Delta S_{R_k} = \log d_{R_k} - \log d_{R_{k-1}}. \tag{4.5.43}$$

The parameters ΔS_{B_k} and ΔS_{R_k} correspond to the change in thermal entropy across the k^{th} time step. This means that, for each $k \neq \hat{k}$, the parameters $a_\pm$ are some $\mathcal{O}(1)$ numbers related to the thermodynamic irreversibility of evaporation. The upshot of all this is that we can approximate the parameter λ (4.5.38) as

$$\frac{\lambda}{d_{\text{code}}} \approx \frac{1 + a_- \frac{d_{R_{\hat{k}-1}}}{d_{B_{\hat{k}-1}}}}{a_+ \frac{d_{R_K}}{d_{B_K}}}. \tag{4.5.44}$$

It is instructive to consider this formula in two different limits:

1. If $\ell_{\hat{k}}$ fell in before the Page time

$$\lambda \approx \frac{1}{a_+} \frac{d_{B_K} d_{\text{code}}}{d_{R_K}}. \tag{4.5.45}$$

This matches the result we got in the simple model (4.5.10), up to an unimportant $\mathcal{O}(1)$ constant. Let $S_{\text{code}} = \log d_{\text{code}}$, then, with the usual identifications of parameters, this means that reconstruction works when

$$S_{\text{rad}} \geq S_{\text{BH}} + S_{\text{code}}. \tag{4.5.46}$$

2. If $\ell_{\hat{k}}$ fell in after the Page time

$$\lambda \approx \frac{a_-}{a_+} \frac{d_{B_K}}{d_{B_{\hat{k}-1}}} \frac{d_{\text{code}}}{d_{R_{\{K,\ldots,\hat{k}\}}}}. \tag{4.5.47}$$

With the usual identifications of parameters, this means that reconstruction works when

$$S_{\text{rad}}^{\text{after diary}} \geq S_{\text{BH}} + S_{\text{code}} - S_{\text{BH}}^{\text{before diary}}. \tag{4.5.48}$$

These twin results are known as the Hayden-Preskill decoding criterion [47]. The delay for reconstructing operators on large code spaces in line with expectations [47, 82, 94]. The conditions for reconstruction (4.5.46) and (4.5.48) also match precisely with what we found in JT gravity using the QES prescription, in section 4.2.3. To consider the effect that thermodynamic irreversibility has on reconstruction in the case that $\ell_{\hat{k}}$ fell in after the Page time, it's useful to rewrite the condition for reconstruction to work (4.5.48) as

$$S_{\text{rad}}^{\text{after diary}} \geq S_{\text{BH}} - S_{\text{BH}}^{\text{after diary}} + \Delta S_{\text{BH}} + S_{\text{code}}. \tag{4.5.49}$$

From the GSL,

$$S_{\text{BH}} - S_{\text{BH}}^{\text{after diary}} \geq 0, \qquad S_{\text{BH}} \geq S_{\text{code}}. \tag{4.5.50}$$

In the reversible limit, both inequalities are saturated and the condition for reconstruction to work (4.5.49) reduces to $S_{\text{rad}}^{\text{after diary}} \geq S_{\text{code}}$, as expected.

4.5.6 An example

Consider a simple example where at each time step the black hole decreases in size by one qubit, except for the $\hat{k}^{\text{th}}$ time step where it also increases in size by ΔS_{BH} qubits, and ξ qubits are emitted in the radiation, that is,

$$d_{B_k} = \begin{cases} 2^{n_{\hat{k}-1}+\hat{k}-1-k} & \text{for } k < \hat{k} \\ 2^{n_K+K-k} & \text{for } k \geq \hat{k} \end{cases}, \qquad d_{R_k} = 2^{\xi k}, \qquad d_{R_{\{k\}}} = 2^{\xi}, \tag{4.5.51}$$

where $n_k = \log d_{B_k}$ is the number of black hole qubits at the k^{th} time step. The point of considering this simple example is that we can actually compute the sums in (4.5.40), up to exponentially small corrections. We find

$$\frac{\lambda}{d_{\text{code}}} \approx \frac{1 + a_{\hat{k}-1} 2^{\xi(\hat{k}-1)-n_{\hat{k}-1}}}{a_{K-\hat{k}+1} 2^{\xi K - n_K}}, \tag{4.5.52}$$

with

$$a_k = \frac{1-2^{-2\xi}}{1-2^{-(1+\xi)}} \left(1-2^{-(1+\xi)k}\right). \tag{4.5.53}$$

In this model, the Page time is

$$k - 1 = \frac{n_0}{\xi}, \tag{4.5.54}$$

where n_0 is the initial number of black hole qubits. If $\ell_{\hat{k}}$ fell in after the Page time, the condition for reconstruction to work is

$$K - \hat{k} + 1 \geq \frac{\Delta S_{\text{BH}} + S_{\text{code}}}{1+\xi}. \tag{4.5.55}$$

4.6 Conclusion and outlook

We have considered how information is recovered in a model of an evaporating black hole in JT gravity, where the backreaction problem for an infalling object is exactly solvable. In the case that we have access to all the Hawking radiation since the formation of the black hole, as discussed in section 4.2.3, we have seen how the QES prescription leads to information recovery in the way anticipated by Hayden and Preskill in their pioneering work [47]. The islands in the stream picture that we developed allowed us to show that we can also recover the information from the radiation emitted after the diary falls in, as discussed in section 4.2.4. The fact that the information in the diary can be recovered in different subsets of the Hawking radiation can be seen as a consequence of the relationship between entanglement wedge reconstruction and quantum error correction [30, 31].

In addition, we have also considered corrections to the entropy which are logarithmic in S_{BH} and the jump ΔS_{SH} in the entropy as the diary is thrown in. These terms are related to the scrambling time. In addition to the scrambling time delay, we have seen how there is an additional delay due to thermodynamic irreversibility (4.2.11) and also the entropy of the diary, which means the black hole doesn't quite behave as a mirror anymore. Using the Python's lunch conjecture in section 4.4 we have seen how irreversibility is also responsible for increasing the complexity of decoding.

In section 4.5 we studied the problem of information recovery in a simple toy model. More specifically, we considered the problem of reconstructing operators acting on the

left-moving ingoing modes. We were able to reproduce the results we found in JT gravity including the salient features of irreversibility and the delay related to the size of the code subspace. For simplicity, we considered a code subspace where the right-moving interior modes r and the right-moving exterior modes R, that is, the Hawking radiation, were maximally entangled. It would be interesting to understand how changing this state affects reconstruction. A simple case study to consider could be a state that is a mixture of the maximally mixed state and some pure state. For this case we expect there to be a more subtle condition for reconstruction to work due to the incompressibility of the state, as discussed in [104, 105].

Chapter 5

Ephemeral islands and BCFT channels

This chapter primarily consists of a paper [106] written in collaboration with Timothy J. Hollowood, S. Prem Kumar, and Andrea Legramandi. Section 5.6 contains unpublished work.

5.1 Introduction

From a holographic perspective, JT gravity coupled to an auxiliary reservoir with conformal matter has a dual description as a quantum mechanical (QM) system coupled to an auxiliary reservoir with the same conformal matter. In an appropriate infrared limit, boundary conformal field theory (BCFT) offers a powerful and tractable framework to examine the microscopic dual of the JT model. In particular, it can offer insight into how entropies in the microscopic dual are captured by the quantum extremal surface prescription in the bulk semiclassical picture.

In this work, the setup involves two copies of the dual QM system coupled to a reservoir, labelled left L and right R, in the thermofield double (TFD) state

$$|\Psi_{\rm TFD}\rangle = \frac{1}{\sqrt{Z}} \sum_n e^{-\beta E_n/2} \, |\bar{E}_n\rangle_L \otimes |E_n\rangle_R \,, \tag{5.1.1}$$

where $|E_n\rangle$ is an energy eigenstate of the right system with energy E_n, $|\bar{E}_n\rangle$ is its time-reversal conjugate in the left system, and Z is the partition function of the left or right system at inverse temperature β. This is expected to be dual to an eternal black hole in the JT model with matter in the Hartle-Hawking state [40].[1] There is a version of

[1] See [107] for another perspective.

the black hole information problem in this setup [56].[2] The thermofield double state is invariant if we evolve forwards on one side and backwards on the other, that is if we evolve with $H_R - H_L$, where H_L and H_R denote the Hamiltonians of the left and right systems, respectively. In the bulk, this corresponds to the fact that the eternal black hole geometry has a boost isometry. However, the thermofield double state is not invariant if evolve forwards on both sides. Tracking the time evolution of the entropy $S(A_L \cup A_R)$ of the left and right reservoirs, A_L and A_R, leads to a potential puzzle. For early times the entropy of quantum fields in $A_L \cup A_R$ on the eternal black hole background grows linearly with time,[3]

$$S_{\text{no island}} \approx \frac{2\pi c}{3\beta} t, \tag{5.1.2}$$

where c is the central charge of the matter quantum field theory which is assumed to be conformal. This leads to a potential puzzle since the entropy of $A_L \cup A_R$ should be bounded above by the thermal entropy of the two QM systems, which is given by twice the Bekenstein-Hawking entropy. However, the island formula indicates that the potential puzzle is resolved by an island (see figure 5.1), which happens to lie just outside the horizon, and has generalised entropy

$$S_{\text{island}} \approx 2S_{\text{BH}}. \tag{5.1.3}$$

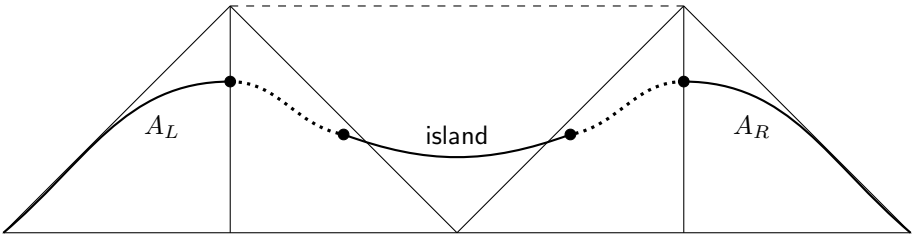


Figure 5.1: A version of the information problem for the eternal black hole. For early times the no island contribution gives an entropy which grows linearly in time. For late times an island configuration, with endpoints just outside the horizon, gives an entropy which is twice the Bekenstein-Hawking entropy.

In this work we study a variant of this setup, where the subregions A_L and A_R are take to be subsets of the left and right reservoirs (see figure 5.2). More specifically, we will be interested in a certain infrared limit, where the two copies of the QM coupled to a reservoir can be replaced by two copies of a BCFT on a half space. The main aims of this work are to use the two complementary descriptions to demonstrate the following:

[2] This differs from Maldacena's version of the information problem for the eternal black hole [40], which considers the late time behaviour of correlation functions and doesn't involve coupling to an auxiliary reservoir. In the next chapter we discuss a problem closely related to this version of the problem.

[3] This formula is valid for times t much larger than the thermal scale.

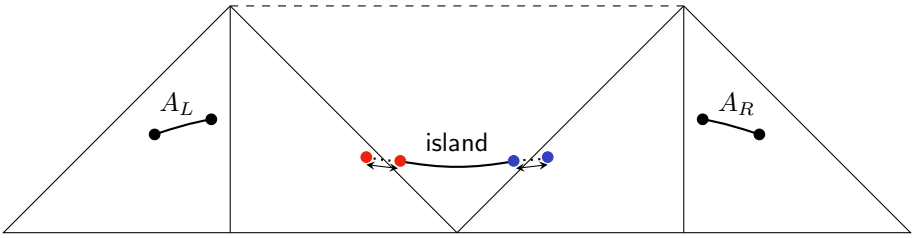


Figure 5.2: The setup considered in this paper. The entropy $S(A_L \cup A_R)$ of the union of two subregions A_L and A_R in the reservoir can receive contributions from an island which, over time, can emerge from behind the horizon and then disappear again behind it.

1. In the high temperature limit, entropies of subregions in BCFT exhibit a nontrivial but universal time evolution across distinct temporal regimes. A caveat here is that this behaviour is really only universal amongst BCFTs which admit an effective quasiparticle description. We comment more on this below.

2. The same behaviour emerges in JT gravity through a nontrivial competition between an empty island an nonempty island. In particular it is possible for a nonempty island to be *ephemeral*, appearing for intermediate times but eventually giving way to the empty island at late times.

3. The distinct temporal regimes in the evolution of the entropy are governed by certain singularities, either singularities in the operator product expansion (OPE) or lightcone singularities[4], in twist operator correlation functions used to compute the Rényi entropy. Singularities in BCFT twist correlators which give rise to disconnected contributions, that is, involving the one point functions of twist operators, get identified with corresponding singularities in the bulk twist correlators which involve island contributions. The island configurations can emerge from behind the horizon and then disappear again behind it.

An example of the time evolution of the entropy, for a symmetric choice of A_L and A_R, is shown in figure 5.3. For early times the entropy increases linearly as modes enter (exit) the interval in one copy of the reservoir and their respective purifiers simultaneously exit (enter) the interval in the second copy of the reservoir. The linear growth then slows and eventually dips before rising again and reaches a plateau at late times. The dip is controlled by a lightcone singularity in the BCFT and an evolving island contribution in the JT model. In the quasiparticle picture it is due to modes reflecting off the boundary in either copy of the reservoirs and entering $A_L \cup A_R$ along with their respective purifiers. In the JT model the dip appears when modes in the reservoir $A_L \cup A_R$ and their purifiers enter the island.

[4]These singularities are intrinsically Lorentzian as they involve an operator hitting the lightcone of two other operators.

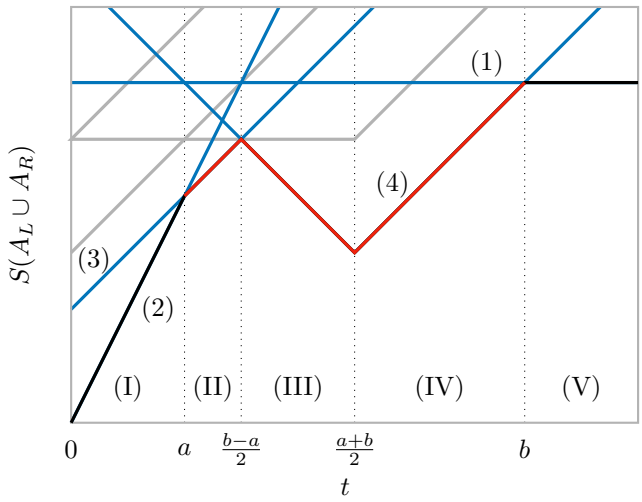


Figure 5.3: BCFT channels for the entropy $S(A_L \cup A_R)$ with $A_L = A_R = [a, b]$ and $a = 3$, $b = 12$ and $\beta = 0.1$. Only the channels shown in blue actually compete, the ones in grey are always subdominant. In the JT model the red portion of the curve corresponds to an island configuration.

A quantitative match between the JT model and BCFT occurs when the characteristic length and time scales are taken to be much larger than not only the thermal scale β but also the scrambling time

$$t_{\text{scr}} = \frac{\beta}{2\pi} \log \frac{S_{\text{BH}} - S_0}{c}. \tag{5.1.4}$$

The scrambling time governs how long one has to wait to recover information which has been thrown into an old black hole. But in the limit where the characteristic scales are taken to sufficiently large, the scrambling time scale is effectively washed away and the black hole behaves like an information mirror, in a similar way to the boundary in BCFT. In this limit we can identify the boundary entropy $\log g_b$ with the Bekenstein-Hawking entropy of the black hole. If we further impose the characteristic scales to be much larger than k then the system matches the behaviour of the free fermion BCFT which has vanishing boundary entropy.

While we use the term *information mirror*, it's crucial to note that unlike a boundary in BCFT, which reflects information coherently, a black hole does not—even when the scrambling time is negligible. However, for small amounts of information, a black hole can act like a mirror if one is just interested in the time scale for recovery. To recover larger amounts of information there will be a delay, unlike for a mirror, due to state dependence, as discussed in chapter 4. Generally, understanding when information is recoverable from a partially evaporated black hole is subtle and perhaps best understood through the framework of quantum error correction.

It's important to emphasise that we focus on CFTs with an effective quasiparticle description [108]. This includes free CFTs, rational CFTs, and in general CFTs whose

high energy density of states is dominated by conserved currents. It's only for these theories where the dip in the entropy, caused by a lightcone singularity, is actually present. In section 5.6 we verify the absence of the dip in large c holographic BCFTs using Takayanagi's formula for computing the entropy holographically [109]. Also, while this work was in preparation the paper [110] appeared which has some overlap with the questions addressed in this paper. In contrast to our setup, where the bulk matter theory is taken be a collection of free fermions, they assume the bulk matter theory is itself holographic and find no dip in the entropy.

The plan of this chapter is as follows. We start by very briefly reviewing the setup and some notation in section 5.2. In section 5.3 we show how the high-temperature limit of the entropy of two intervals in BCFT is controlled by certain singularities in the correlation function of twist fields. In section 5.4 we show how this behaviour is reproduced in the JT model, highlighting the connection between bulk and boundary channels. In section 5.5 we explain how the entropy in BCFT follows from a quasi-particle picture while the entropy in the JT model follows from tracing modes and the basic entanglement structure across the horizon. In section 5.6 we compute the entropy of two intervals in a holographic BCFT using Takayanagi's formula and verify the absence of the dip. We end with some conclusions and outlook for future work in section 5.7.

5.2 Preliminaries

We consider two copies of the BCFT on the half space $x \geq 0$ in the thermofield double state, as shown in figure 5.4. The focus will be on tracking the time evolution of an interval $A_L \cup A_R$ where $A_L = [a_1, b_1]$ is an interval in the left reservoir and $A_R = [a_2, b_2]$ is an interval in the right bath. For simplicity we will initially focus on the symmetric case: $a \coloneqq a_1 = a_2$ and $b \coloneqq b_1 = b_2$. Left-right correlators in the thermofield double state can be computed by analytic continuation in t_L from right-right thermal correlators.

In semiclassical picture we have JT gravity with a bulk matter conformal field theory consisting of c massless Dirac fermions in the Hartle-Hawking state. This is the vacuum associated to the Kruskal frame which has null coordinates $w^{\pm}$ that cover the entire spacetime. These are related to null coordinates $x^{\pm}_{L,R} = t_{L,R} \pm x_{L,R}$, defined in left and right exterior regions, by

$$w^{\pm} = \pm e^{\pm 2\pi x^{\pm}_{L,R}/\beta}. \tag{5.2.1}$$

The endpoints of the intervals A_L and A_R will be indexed by $2_L, 1_L, 1_R$, and 2_R, ordered along a Cauchy surface from left to right, as shown in figure (5.5). To evolve forward in time on both sides by time t we set

$$t_L = -t + \frac{\mathrm{i}\beta}{2}, \qquad t_R = t. \tag{5.2.2}$$

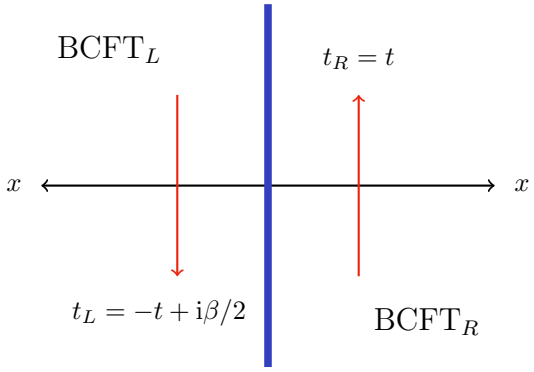


Figure 5.4: Two copies of the BCFT on the half space $x \geq 0$. Correlators in the thermofield double are related to ordinary thermal correlators by analytic continuation in time.

The metric and dilaton profile are (see section 2.4):

$$\mathrm{d}s^2 = -\frac{4\mathrm{d}w^+\mathrm{d}w^-}{(1+w^+w^-)^2}, \qquad \phi = \frac{2\pi\phi_r}{\beta}\frac{1-w^+w^-}{1+w^+w^-}. \tag{5.2.3}$$

For this chapter we use a different notation than in the previous chapters. To compare notation we should identify $w^- = U, w^+ = V, x^- = u$, and $x^+ = v$.

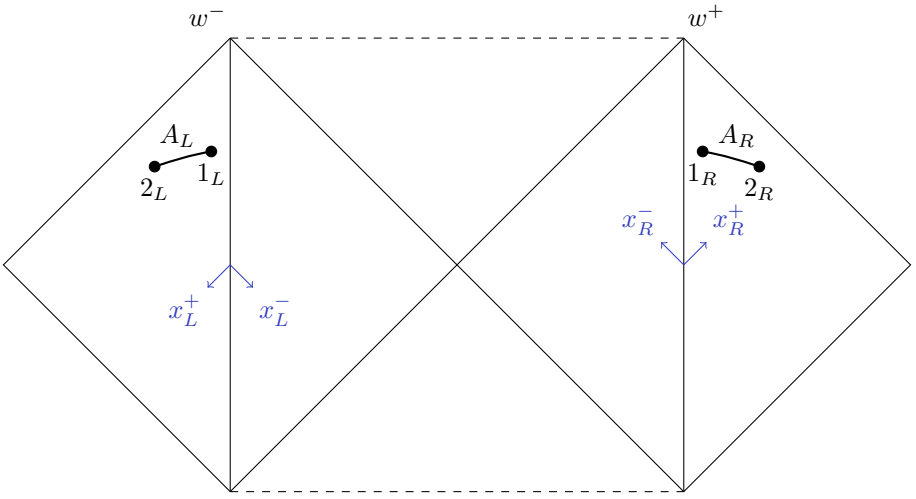


Figure 5.5: The Penrose diagram of the eternal black hole.

5.3 BCFT

In this section we compute the time evolution of the entropy of the subregion $A_L \cup A_R$ in two copies of a BCFT in the thermofield double state. The strategy is to relate the

Rényi entropy to a correlation function of *twist fields*. Using Cardy's doubling trick [111] we can then relate correlation functions of twist fields on the half plane to correlation functions on the plane. We begin this section with some background material on twist fields and Cardy's doubling trick.

5.3.1 Replica trick and twist fields

For a two-dimensional quantum field theory on the plane, the reduced density matrix ρ_A of the vacuum state on an interval $A = [0, L]$ can be computed by a path integral on the plane with a cut along A. $\mathrm{Tr}\rho_A^n$ can then be computed by a path integral Z_n on an n-sheeted Riemann surface, constructed by cyclically gluing n copies of the cut plane along the cuts. This gives

$$\mathrm{Tr}\rho_A^n = \frac{Z_n}{Z_1^n}, \tag{5.3.1}$$

where dividing by the path integral Z_1 on the plane ensures $\mathrm{Tr}\rho_A = 1$. It is convenient to view the path integral of the QFT on the n-sheeted Riemann surface as the path integral of n copies of the QFT on the plane, with the structure of the Riemann surface implemented through a nontrivial monodromy of the fields around the end points of the interval. The nontrivial monodromy can be implemented by inserting a *twist* $\mathcal{T}(0)$ and *antitwist field* $\bar{\mathcal{T}}(L)$ at the end points of the interval, so that

$$\mathrm{Tr}\rho_A^n = \langle \mathcal{T}(0)\bar{\mathcal{T}}(L)\rangle. \tag{5.3.2}$$

In a CFT the twist and antitwist fields are scalar primary fields with dimension[5]

$$\Delta = \frac{c}{12}\left(n - \frac{1}{n}\right), \tag{5.3.3}$$

where c is the central charge of one copy of the CFT. In a similar way, the Rényi entropy of a union N intervals is determined by a $2N$-point function of twist and antitwist fields. This implies that the Rényi entropy is only universal for the case of a single interval, as it is determined by the two-point function of primary fields, whose form is fixed by conformal invariance. For the single interval A,

$$\mathrm{Tr}\rho_A^n = \left(\frac{\epsilon}{L}\right)^{\frac{c}{6}\left(n-\frac{1}{n}\right)}, \tag{5.3.4}$$

where ϵ is a normalisation which can interpreted as a ultraviolet cutoff. Analytically continuing the Rényi entropy to $n = 1$ gives the von Neumann entropy

$$S(A) = \frac{c}{3}\log\frac{L}{\epsilon}. \tag{5.3.5}$$

[5]Under the state-operator correspondence a twist field inserted at the origin $\mathcal{T}(0)$ is mapped to a state $|\mathcal{T}\rangle$ of the n copy CFT on a cylinder of circumference 2π with energy $E = \Delta - \frac{nc}{12}$, where Δ is the conformal dimension of $\mathcal{T}$. Instead viewing this as the path integral of one copy of the CFT on the n-sheeted cover of the cylinder with circumference n times bigger gives $E = -\frac{c}{12n}$. Equating both expressions gives $\Delta = \frac{c}{12}\left(n - \frac{1}{n}\right)$.

Twist field correlators in the thermofield double state can be obtained by analytic continuation of thermal correlators, which in turn can be obtained from vacuum correlators by mapping the plane to the thermal cylinder.

5.3.2 Cardy's doubling trick

The standard model for a CFT on a two dimensional manifold with boundary is the upper half plane (UHP) $\mathrm{Im}\, z \geq 0$. Correlation functions in more general geometries can be computed by conformally mapping back to the UHP. This section is intended to be a lightning review focusing on the minimum needed to understand the computations in following sections. More complete reviews on BCFT are [112–114].

The presence of the boundary breaks conformal symmetry, most obviously translations perpendicular to the boundary. To preserve some form of conformal symmetry, conformal transformations which map the UHP to itself and respect the boundary should remain a symmetry. For an infinitesimal transformation $(x, y) \mapsto (x, y) + \epsilon(x, y)$ this means $\epsilon_y(x, 0) = 0$. Now suppose ϵ is conformal in a semi-disc D centred around a point $w \in$ UHP near the real axis but arbitrary elsewhere in a compact region $C \subset$ UHP. Since the action of an arbitrary infinitesimal coordinate transformation is implemented by the stress tensor,

$$\begin{aligned}\delta\phi(w,\bar{w}) &= -\int_{C-D} \partial_a \epsilon_b(x,y) T^{ab}(x,y)\phi(w,\bar{w}) \\ &= -\int_{\partial(C-D)} n_a \epsilon_b(x,y) T^{ab}(x,y)\phi(w,\bar{w}) + \int_{C-D} \epsilon_b(x,y)\partial_a T^{ab}(x,y)\phi(w,\bar{w}).\end{aligned} \tag{5.3.6}$$

Since the LHS is independent of the values of ϵ outside D we must have $\partial_a T^{ab}$, which is just conservation of stress energy, but we also get another constraint from the portion of the first term which is integrated along a segment $[a, b]$ of the real axis

$$\delta\phi(w,\bar{w}) \supset -\int_a^b \epsilon_x(x,0) T^{xy}(x,0)\phi(w,\bar{w}). \tag{5.3.7}$$

Demanding this be independent of ϵ implies the conformal boundary condition $T_{xy}(x, 0) = 0$, which has the interpretation of the absence of energy flow across the boundary [111]. In complex coordinates

$$T(z) = \bar{T}(\bar{z}) \quad \text{for } z = \bar{z}. \tag{5.3.8}$$

This relation links the holomorphic and antiholomorphic components of the stress tensor which were independent on the full complex plane. Notice that the constraint (5.3.8) does not uniquely fix the boundary conditions but merely selects a class of boundary conditions compatible with conformal invariance. It turns out that there are a discrete set of compatible boundary conditions which we will label by an index b.

The conformal boundary condition (5.3.8) implies by analytic continuation that $T(z) = \bar{T}(\bar{z})$ for all $z \in \mathbb{C}$. The analytic continuation of T from the UHP to the full

complex plane has the same algebraic properties as one chiral half of the stress tensor on the full complex plane. To understand the constraints conformal invariance places on correlation functions on the UHP, consider a correlation function of a collection of primary fields $\phi_k(w_k, \bar{w}_k)$ with weights $(h_k, \bar{h}_k)$ together with an insertion of the holomorphic component of the stress tensor. This leads to the conformal Ward identity on the UHP,

$$\left\langle T(z) \prod_{k=1}^{N} \phi_k(w_k, \bar{w}_k) \right\rangle_{\text{UHP}} = \sum_{k=1}^{N} \left[\frac{h_k}{(z-w_k)^2} + \frac{\partial_{w_k}}{z-w_k} + \frac{\bar{h}_k}{(\bar{z}-\bar{w}_k)^2} + \frac{\partial_{\bar{w}_k}}{\bar{z}-\bar{w}_k} \right] \left\langle \prod_{k=1}^{N} \phi_k(w_k, \bar{w}_k) \right\rangle_{\text{UHP}} . \tag{5.3.9}$$

This should be compared with the conformal Ward identity on the full complex plane for which the differential operator in square brackets has no antiholomorphic dependence. From this observation, it follows that correlation functions in a BCFT on the UHP are constrained by the conformal Ward identities to have the same functional form as a correlation function with twice as many chiral fields in a CFT on the full complex plane,

$$\left\langle \prod_{k=1}^{N} \phi_k(w_k, \bar{w}_k) \right\rangle_{\text{UHP}} \quad \Leftrightarrow \quad \left\langle \prod_{k=1}^{N} \phi_k(w_k) \prod_{k=1}^{N} \bar{\phi}_k(\bar{w}_k) \right\rangle_{\mathbb{C}} , \tag{5.3.10}$$

where $\phi_k(w_k)$ is a chiral field with weights $(h_k, 0)$ and $\bar{\phi}_k(\bar{w}_k)$ is a chiral field with weights $(0, \bar{h}_k)$ placed at an image point $\bar{w}_k$. This is known as Cardy's doubling trick. The simplest consequence of the doubling trick is that one-point functions in a BCFT can be nonvanishing. In particular, the entropy of an interval $A = [0, L]$ including the boundary in a BCFT is computed by the one-point function of a twist field, which by the doubling trick, is determined by a two-point function,

$$\langle \mathcal{T}(L) \rangle_{\text{UHP}} = g_b^{1-n} \left(\frac{\epsilon}{2L} \right)^{\frac{c}{12}\left(n - \frac{1}{n}\right)} . \tag{5.3.11}$$

Analytically continuing the Rényi entropy to $n = 1$ gives

$$S(A) = \frac{c}{6} \log \frac{2L}{\epsilon} + \log g_b . \tag{5.3.12}$$

The quantity $\log g_b$ is known as the boundary entropy [115], which depends on the boundary condition b.

5.3.3 Half lines

As a warm up, we first consider the case that the subregions A_L and A_R are the half lines $[a_1, \infty]$ and $[a_2, \infty]$. This situation was analysed from the perspective of the eternal black hole in [12, 56], and its higher dimensional realisations in [116]. A variant of this problem, where the BCFT is taken to live on a strip instead of a half space, was considered in [117] for a large c holographic BCFT. This example will illustrate how the exact result at high temperatures is essentially determined by a competition between different channels for the OPE. We will first focus on free fermion BCFT, for which exact results are known, and then argue that at high temperature the features are universal for BCFTs that admit a quasiparticle description.

For free fermions the formula for the Rényi entropy of a disjoint union of an arbitrary number of intervals is known [79]. Viewing the free fermion BCFT as a chiral fermion theory on the full complex plane gives the following expression for the Rényi entropy of $A_L \cup A_R$, which is governed by a four-point function of twist fields,

$$S^{(n)}(A_L \cup A_R) = \frac{c(n+1)}{12n}\left(\log\sinh\tfrac{\pi}{\beta}(x^+_{1_L} - x^+_{1_R})\sinh\tfrac{\pi}{\beta}(x^-_{1_L} - x^-_{1_R}) + \log\eta\right), \quad (5.3.13)$$

where

$$\eta = \frac{\sinh\frac{\pi}{\beta}(x^+_{1_L} - x^-_{1_L})\sinh\frac{\pi}{\beta}(x^+_{1_R} - x^-_{1_R})}{\sinh\frac{\pi}{\beta}(x^-_{1_R} - x^+_{1_L})\sinh\frac{\pi}{\beta}(x^-_{1_L} - x^+_{1_R})}. \quad (5.3.14)$$

is the cross ratio. Here we have suppressed the additive logarithmic dependence on the ultraviolet cutoff. For a general BCFT there will also be a contribution from the boundary entropy $\log g_b$, but for the free fermion theory it is vanishes. We will focus primarily on the von Neumann entropy, which is determined by analytically continuing to $n = 1$.

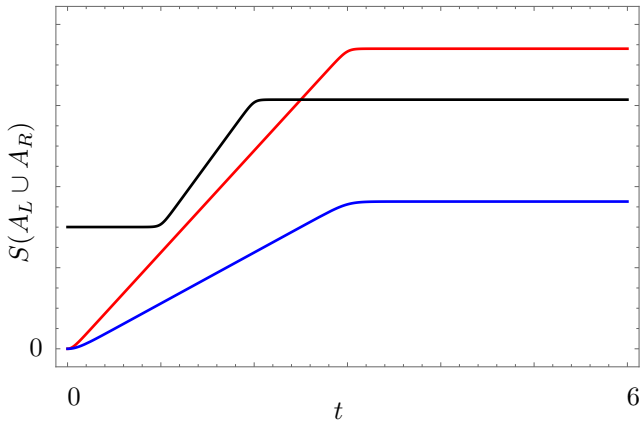


Figure 5.6: The entropy for free fermion BCFT with $A_L = [a_1, \infty]$ and $A_R = [a_2, \infty]$. The different colour curves correspond to the cases: $a_1 = a_2 = 3$, $\beta = 1$ (blue), $a_1 = a_2 = 3$, $\beta = 0.5$ (red), and $a_1 = 1$, $a_2 = 3$, $\beta = 0.4$ (black).

The entropy for the symmetric case $a = a_1 = a_2$ in free fermion BCFT is shown for two choices of parameters in figure 5.6 as the blue and red curves. As the entropy

smoothly interpolates between early time linear growth and late time saturation, the cross ratio η interpolates between 1 and 0. The transition between the two regimes becomes sharper as temperature is increased. The early time linear growth is governed by a connected channel (see the left panel of figure 5.7) with $\eta = 1$ for which the entropy reduces to single interval entropy for the region between A_L and A_R.

The late time saturation is instead due to a disconnected channel with $\eta = 0$ which gives the product of one-point functions of the twist and antitwist fields

$$\left\langle \mathcal{T}(x^{\pm}_{1_L})\bar{\mathcal{T}}(x^{\pm}_{1_R})\right\rangle_{\text{BCFT}} \overset{\eta\to 0}{=} \left\langle \mathcal{T}(x^{\pm}_{1_L})\right\rangle \left\langle {}_{\beta}\bar{\mathcal{T}}(x^{\pm}_{1_R})\right\rangle_{\text{BCFT}}. \tag{5.3.15}$$

The behaviour of the correlator closely follows the expectation for large c CFTs where the two different regimes can be viewed as distinct saddle point approximations to the large c conformal blocks [118, 119].[6] A useful lesson from this example is that the corrections to the two asymptotic behaviours at early and late times are exponentially small in the thermodynamic limit. For the symmetric situation $a = a_1 = a_2$,

$$\eta \approx \begin{cases} 1 - e^{-\frac{4\pi}{\beta}(a-t)} - e^{-\frac{4\pi}{\beta}(a+t)} - 2e^{-\frac{4\pi}{\beta}a}, & t < a, \\ e^{-\frac{4\pi}{\beta}(t-a)}\left(1 - e^{-\frac{4\pi}{\beta}(t-a)} - e^{-\frac{4\pi}{\beta}(a+t)} - 2e^{-\frac{4\pi}{\beta}a}\right), & t > a. \end{cases} \tag{5.3.16}$$

The high temperature thermodynamic limit effectively erases short range correlations and gives a universal result for entropies, and unitarity makes its appearance through competition between two channels.

In the high temperature limit, where free fermion and large c BCFT appear to coincide in this example, the formula for the entropy can be written in terms of the thermodynamic entropy of the left- and right-moving radiation in an interval of null coordinate $[x_1^{\pm}, x_2^{\pm}]$, that is,

$$S_{\text{rad}}(x_1^{\pm}, x_2^{\pm}) := \frac{\pi c}{6\beta}|x_1^{\pm} - x_2^{\pm}|. \tag{5.3.17}$$

The entropy for the interval $A_L \cup A_R$ is then

$$\begin{aligned} S(A_L \cup A_R) = \min\big\{ & S_{\text{rad}}(x^+_{1_L}, x^+_{1_R}) + S_{\text{rad}}(x^-_{1_L}, x^-_{1_R}), \\ & S_{\text{rad}}(x^+_{1_L}, x^-_{1_L}) + S_{\text{rad}}(x^+_{1_R}, x^-_{1_R}) + 2\log g_b\big\}. \end{aligned} \tag{5.3.18}$$

Here we have allowed for a nonzero boundary entropy g_b and implemented the rule that every one-point function, understood as a contraction between a point and its image, is accompanied by the boundary entropy $\log g_b$. We can also write this as

$$\begin{aligned} S(A_L \cup A_R) = \frac{\pi c}{6\beta} \min\big\{ & |2t - (a_2 - a_1)| + |2t + (a_2 - a_1)|, \\ & 2(a_1 + a_2) + \tfrac{12\beta}{\pi c}\log g_b\big\}. \end{aligned} \tag{5.3.19}$$

[6] An argument for this follows from modular invariance which suggests that correlators in the high temperature limit should be determined by the vacuum conformal block. The saddle point interpretation then arises as a consequence of the thermodynamic limit instead of large c.

When $\log g_b = 0$ this reproduces the curve for free fermion curve BCFT in figure 5.6. The two channels exchange dominance at the crossover time, which will be identified with the Page time in the bulk semiclassical picture,

$$t_{\text{Page}} = \frac{a_1 + a_2}{2} + \frac{3\beta}{\pi c} \log g_b. \tag{5.3.20}$$

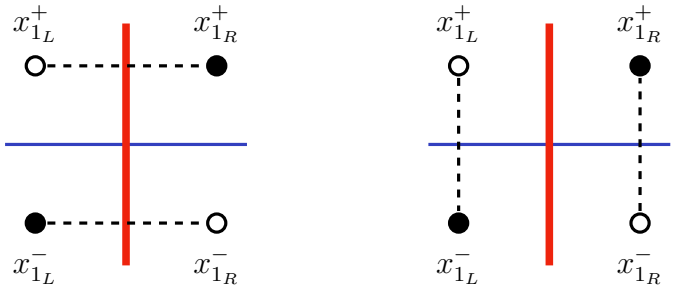


Figure 5.7: A schematic representation of the two distinct channels for the OPE of a twist-antitwist correlator in BCFT. The panel on the left is a connected contribution. The panel on the right is a disconnected contribution given by the product of one-point function of the twist and antitwist fields on the half space. Solid and hollow dots distinguish twist and antitwist fields.

5.3.4 Two intervals

We now turn to the case that A_L and A_R are the intervals $[a_1, b_1]$ and $[a_2, b_2]$, as described in section 5.2. In particular we assume $a_1, a_2 > 0$, that is, the intervals don't include the boundary. The Rényi entropy of $A_L \cup A_R$ is determined by the analytic continuation of the four-point function of twist fields on half of the thermal cylinder

$$S^{(n)}(A_L \cup A_R) = \left\langle \mathcal{T}(x^{\pm}_{1_L}) \bar{\mathcal{T}}(x^{\pm}_{2_L}) \mathcal{T}(x^{\pm}_{1_R}) \bar{\mathcal{T}}(x^{\pm}_{2_R}) \right\rangle_{\text{BCFT}}. \tag{5.3.21}$$

By the doubling trick this is related to an eight-point function of chiral fields on the full thermal cylinder. There are now several OPE channels, which correspond to pairs of twist and antitwist fields approaching each other.

For free fermions the BCFT the Rényi entropy can be determined using [79] together with the doubling trick. As usual the von Neumann entropy is then determined by the analytic continuation to $n = 1$. The result is conveniently written in terms of the mutual information,

$$S(A_L \cup A_R) = S(A_L) + S(A_R) - I(A_L : A_R). \tag{5.3.22}$$

The entropies of A_L is

$$S(A_L) = \frac{c}{6} \log \frac{\text{sh}\frac{\pi}{\beta}(x^+_{2_L} - x^+_{1_L})\text{sh}\frac{\pi}{\beta}(x^-_{2_L} - x^-_{1_L})\text{sh}\frac{\pi}{\beta}(x^+_{1_L} - x^-_{1_L})\text{sh}\frac{\pi}{\beta}(x^+_{2_L} - x^-_{2_L})}{\text{sh}\frac{\pi}{\beta}(x^-_{2_L} - x^+_{1_L})\text{sh}\frac{\pi}{\beta}(x^-_{1_L} - x^+_{2_L})}, \tag{5.3.23}$$

where we adopted the shorthand sh = sinh. There is a similar expression for the entropy of A_R. The mutual information between A_L and A_R is

$$I(A_L:A_R) = -\frac{c}{6}\log\left\{\frac{\mathrm{sh}\frac{\pi}{\beta}(x^+_{1_L}-x^+_{1_R})\mathrm{sh}\frac{\pi}{\beta}(x^+_{2_L}-x^+_{2_R})\mathrm{sh}\frac{\pi}{\beta}(x^-_{1_L}-x^-_{1_R})\mathrm{sh}\frac{\pi}{\beta}(x^-_{2_L}-x^-_{2_R})}{\mathrm{sh}\frac{\pi}{\beta}(x^+_{1_L}-x^+_{2_R})\mathrm{sh}\frac{\pi}{\beta}(x^+_{2_L}-x^+_{1_R})\mathrm{sh}\frac{\pi}{\beta}(x^-_{1_L}-x^-_{2_R})\mathrm{sh}\frac{\pi}{\beta}(x^-_{2_L}-x^-_{1_R})}\right.$$
$$\left.\times\frac{\mathrm{sh}\frac{\pi}{\beta}(x^+_{1_L}-x^-_{2_R})\mathrm{sh}\frac{\pi}{\beta}(x^+_{2_L}-x^-_{1_R})\mathrm{sh}\frac{\pi}{\beta}(x^+_{1_R}-x^-_{2_L})\mathrm{sh}\frac{\pi}{\beta}(x^+_{2_R}-x^-_{1_L})}{\mathrm{sh}\frac{\pi}{\beta}(x^+_{1_L}-x^-_{1_R})\mathrm{sh}\frac{\pi}{\beta}(x^+_{2_L}-x^-_{2_R})\mathrm{sh}\frac{\pi}{\beta}(x^+_{1_R}-x^-_{1_L})\mathrm{sh}\frac{\pi}{\beta}(x^+_{2_R}-x^-_{2_L})}\right\}. \tag{5.3.24}$$

The entropy of $A_L \cup A_R$ is shown for two choices of parameters in figure 5.8.

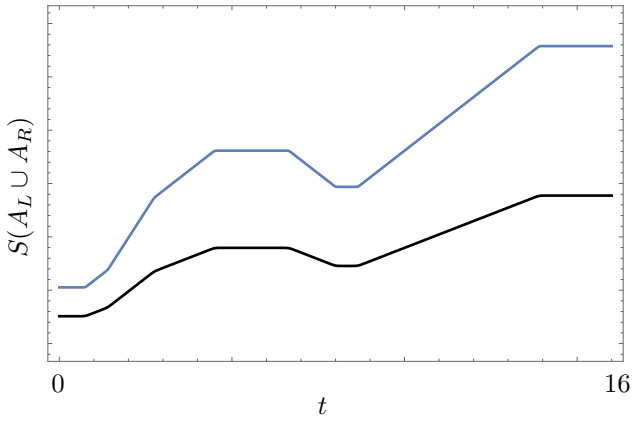


Figure 5.8: The entropy for free fermion BCFT with $A_L = [a_1, b_1]$ and $A_R = [a_2, b_2]$ with $a_1 = 2, b_1 = 12.5, a_2 = 3.5$, and $b_2 = 15.3$. The different colour curves correspond to the cases that $\beta = 0.25$ (blue) and $\beta = 0.5$ (black). The dip in the entropy is clearly present.

5.3.5 OPE channels

As the temperature is increased, the entropy of $A_L \cup A_R$ for free fermion BCFT approaches a piecewise linear function, as seen in figures 5.6 and 5.8. This behaviour at high temperatures can be understood as a competition between different OPE channels for the correlation function of twist and antitwist fields. The channels can be characterised by the distinct ways the twist operators can pair with the antitwist operators. For the case of two intervals there are 4! possible channels. Most of the channels typically have high entropy and so only a small number of channels can actually compete, as shown in figures 5.9 and 5.11. We expect this behaviour to be universal for BCFTs which admit a quasiparticle description. In particular, some of the channels, or more precisely, singularities, that are seen for the free fermion theory are not expected to appear for nonrational and large c holographic BCFTs [108].

We first consider the case of two symmetric intervals, that is, A_L is taken to a copy of A_R with $a = a_1 = a_2$ and $b = b_1 = b_2$. The four relevant channels are shown in 5.9.

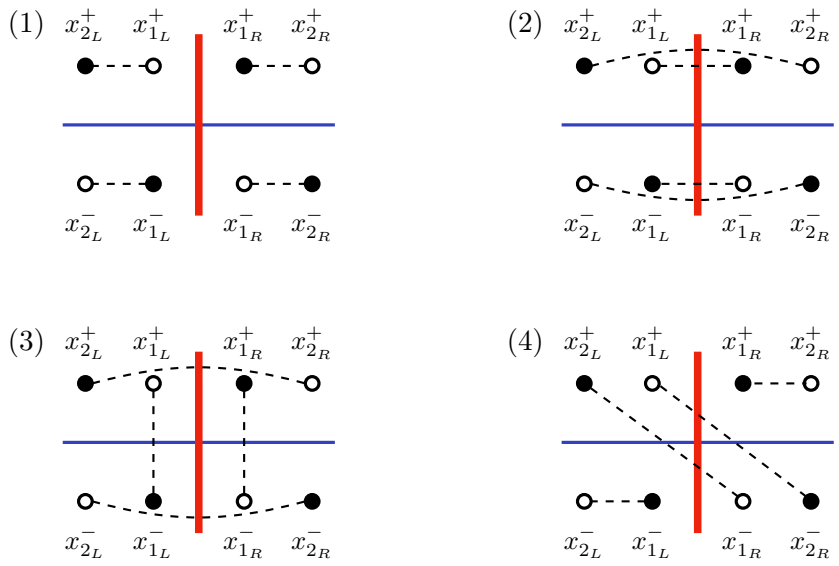


Figure 5.9: OPE channels for the Rényi entropy of two intervals. The third panel is a disconnected contribution which involves one-point functions of the twist and antitwist fields. The fourth panel corresponds to a lightcone singularity. Both the third and fourth panels will be identified with island contributions in the JT model.

At high temperatures the entropy is determined by the piecewise linear function

$$S(A_L \cup A_R) = \frac{2\pi c}{3\beta} \min \Big\{ (b-a\,,\, 2t\,,\, t+a+\tfrac{3\beta}{\pi c}\log g_b\,, \\ \tfrac{1}{2}(b-a) + \big|t - \tfrac{1}{2}(a+b)\big| + \tfrac{3\beta}{\pi c}\log g_b \Big\}\,. \tag{5.3.25}$$

Two of the channels give disconnected contributions which give rise to a boundary entropy term. For the free fermion theory $\log g_b$ should be set to zero. From (5.3.25) it is possible to identify several distinct temporal regimes for the entropy. First consider the case that $b > 3a$. The five distinct temporal regimes are shown in figure 5.3. The early time linear growth (region I) is controlled by a connected channel, shown in panel (2) of figure 5.9. The linear growth then slows (region II) as there is a transition to a partially disconnected channel, shown in panel (3) of figure 5.9, at a time

$$t_{2\to 3} = a + \frac{3\beta}{\pi c}\log g_b. \tag{5.3.26}$$

The linear growth then gives way to a dip before rising again (regions III and IV) as there is a transition to another partially disconnected channel, shown in panel (4) of figure 5.9, at a time

$$t_{3\to 4} = \frac{b-a}{2}. \tag{5.3.27}$$

Actually, since the $x^{\pm}$ coordinates are in different channels, this does not correspond to an OPE limit. Instead it corresponds to an operator hitting the lightcone of two other operators. For the free fermion theory, the entropy attains a minimal value at $t_{\text{dip}} = (a+b)/2$ given by the thermal entropy of either A_L or A_R alone, that is,

$$S(A_L \cup A_R)|_{t=t_{\text{dip}}} = S(A_L) = S(A_R). \tag{5.3.28}$$

After the dip and rise, the entropy eventually reaches a plateau (region V) where the entropy is given by the thermal entropy of either A_L or A_R alone. This happens when there is a transition to a connected channel, shown in panel (1) of figure 5.9, at a time

$$t_{4\to 1} = b - \frac{3\beta}{\pi c}\log g_b. \tag{5.3.29}$$

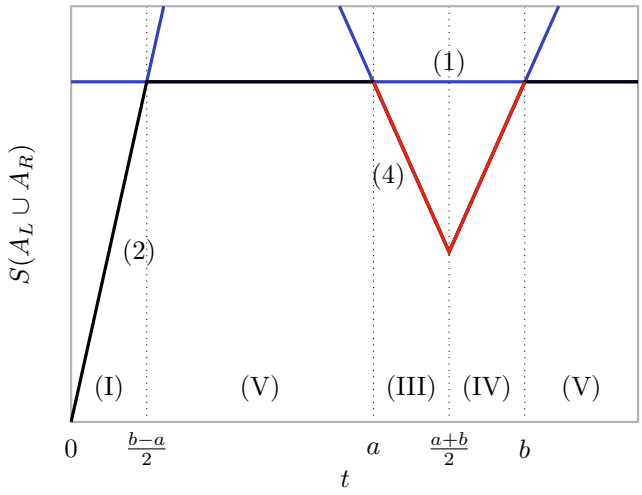


Figure 5.10: The entropy of two small intervals A_L and A_R, that is, with $b < 3a$ for $a = 7$ and $b = 8$. The labels $(1), (2), \ldots$ refer to the channels in figure 5.9 which contribute. Channel (3) does not contribute in this scenario. In the JT model the red portion of the curve corresponds to an island configuration.

For the case that $b < 3a$ there can be different sequence of temporal regimes, as shown in figure 5.10, provided the boundary entropy $\log g_b$ is small enough. Referring to the channel shown in figure 5.9, in this scenario, channel (3) is always subdominant and there is a direct transition from linear growth to a plateau as there is a transition from channel (2) to channel (1) at a time

$$t_{2\to 1} = \frac{b-a}{2}. \tag{5.3.30}$$

However, the plateau is disturbed by a dip and rise as there is an exchange of channels (1) and (4) at a time

$$t_{1\to 4} = a + \frac{3\beta}{\pi c}\log g_b. \tag{5.3.31}$$

For the generic case with two intervals A_L and A_R, that is, A_L is not an identical copy of A_R there are additional channels (figure 5.11) which can contribute. The effect of these additional channels is shown in figure 5.12. The crossover between channels (3) and (4) is now flattened out by the appearance of channel (5). Additionally, the dip from channel (4) is also flattened out. By varying the intervals one can see how various connected and disconnected channels become dominant at intermediate times while the early time growth and late time plateau are always controlled by the connected channels (2) and (1), respectively.

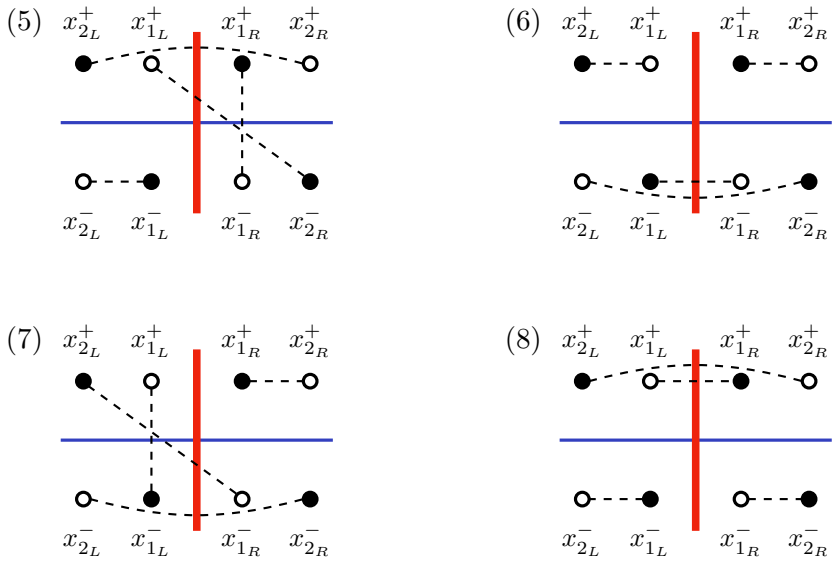


Figure 5.11: Additional OPE channels for the Rényi entropy of two intervals which can contribute in the generic case where A_L is not an identical copy of A_R.

5.4 JT gravity

In this section we compute the time evolution of the entropy of the subregion $A_L \cup A_R$ in JT gravity coupled to free fermions in the Hartle-Hawking state using the QES prescription

$$S(R) = \min_I \operatorname{ext} S_{\text{gen}}(I \cup R) = \min_I \operatorname{ext} \left[\frac{A(\partial I)}{4G} + S_{\text{QFT}}(I \cup R) \right]. \tag{5.4.1}$$

In JT gravity the area A is replaced by the dilaton $\phi_0 + \phi$. In this section we will see how entropy in the JT model is related to the entropy in the BCFT model. In particular, disconnected channels in the BCFT model which receive contributions from the boundary entropy $\log g_b$ will turn out to correspond to configurations with an island.

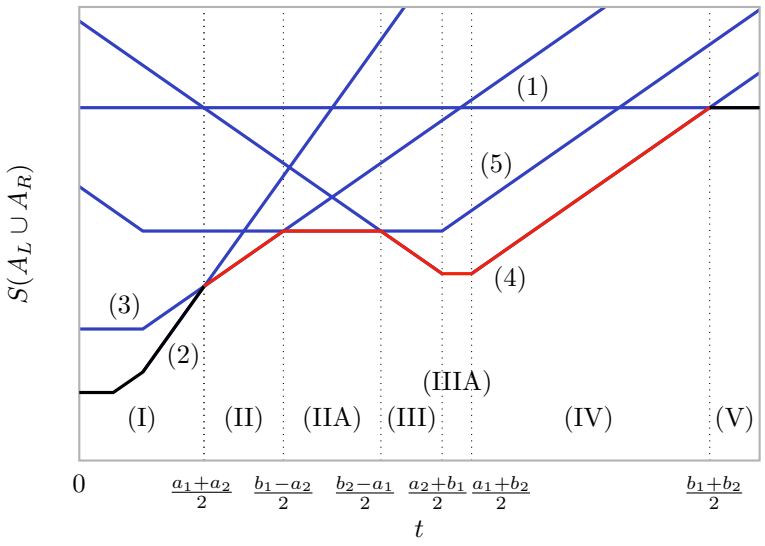


Figure 5.12: The entropy of two intervals $A_L = [a_1, b_2]$ and $A_R = [a_2, b_2]$ with $a_1 = 2, b_1 = 12.5, a_2 = 3.5$, and $b_2 = 15.3$. The competing channels are shown in blue and the labels $(1), (2), \ldots$ refer to the channels in figure 5.9 and 5.11. In the JT model the red portion of the curve corresponds to an island configuration.

On the other hand, connected channels in the BCFT model will turn out to correspond to configurations without an island.

5.4.1 Half lines

As in the previous section, we begin with a warm up, by first considering the case that the subregions A_L and A_R are identical copies of the half line $[a, \infty]$. This situation has been analysed in [56]. At early times, there is no island, and the entropy grows linearly

$$S_{\text{gen}}(A_L \cup A_R) = \frac{2\pi c}{3\beta} t. \tag{5.4.2}$$

Past the Page time, the entropy plateaus as there is a contribution from an island I which has generalised entropy

$$S_{\text{gen}}(I \cup A_L \cup A_R) = 2S_{\text{BH}}. \tag{5.4.3}$$

At very late times the endpoints of the island, that is, the QESs, lie just outside the horizon. Interestingly, a careful analysis reveals the QESs begin their life behind the horizon and then migrate out at an intermediate time scale. The location of the QESs in Kruskal coordinates are

$$w^-_{1_Q} = -\frac{s}{w^+_{1_L}}, \qquad w^+_{1_Q} = -\frac{s}{w^-_{1_L}} - \frac{1}{s} w^-_{1_R} \left(w^+_{1_L}\right)^2, \qquad w^\pm_{2_Q} = w^\mp_{1_Q}, \tag{5.4.4}$$

where $w^{\mp}_{1_L} = w^{\pm}_{1_R} = \pm e^{2\pi(\pm t+a)/\beta}$ are the endpoints of the intervals A_L and A_R in the reservoir and we have defined

$$s \coloneqq \frac{c}{12}\frac{1}{S_{\rm BH}-S_0} = \frac{k\beta}{2\pi} \ll 1. \tag{5.4.5}$$

The QESs always remain close to the horizon, that is, either the left or right future horizon, which correspond to the surfaces $w^- = 0$ and $w^+ = 0$, respectively. The QESs begin their life behind the horizon and exit at the time

$$t_{\rm exit} = 2a + \frac{\beta}{2\pi}\log\frac{c}{12(S_{\rm BH}-S_0)}. \tag{5.4.6}$$

The second term in this expression is the scrambling time which, in the following, will be taken to be small compared to the length scales of the intervals. For large enough intervals, this is larger than the Page time

$$t_{\rm Page} = a + \frac{3\beta}{\pi c}S_{\rm BH}. \tag{5.4.7}$$

The entropy of $A_L \cup A_R$ in the JT model is precisely (5.3.19) if we identify the boundary entropy contribution $\log g_b$ with the Bekenstein-Hawking entropy $S_{\rm BH}$.

5.4.2 Two intervals

We now turn to the main case of interest, where A_L and A_R are identical copies of the interval $[a,b]$. In particular we assume $a > 0$, that is, the intervals lie entirely with the reservoir region. As in the previous section, we begin by analysing the case that $b > 3a$, for which it is useful to separate the analysis into the five temporal regimes

$$\begin{aligned} \text{(I)} &= \{0 < t < a\}, & \text{(II)} &= \{a < t < \tfrac{1}{2}(b-a)\}, \\ \text{(III)} &= \{\tfrac{1}{2}(b-a) < t < \tfrac{1}{2}(a+b)\}, & \text{(IV)} &= \{\tfrac{1}{2}(a+b) < t < b\}, \\ \text{(V)} &= \{b < t\}. \end{aligned} \tag{5.4.8}$$

The main strategy to solve for the QESs will be to exploit the fact that in the high temperature limit the distances between the endpoints of the intervals A_L and A_R and the QESs is exponentially large in the Kruskal coordinates $w^{\pm}$. We begin by studying the behaviour of the generalised entropy of the no island contribution. In regions (I) and (II), that is for $t < \frac{1}{2}(b-a)$, the endpoints of the intervals A_L and A_R are ordered as

$$x^+_{1_L} < x^+_{1_R} < x^+_{2_L} < x^+_{2_R}, \qquad x^-_{2_L} < x^-_{2_R} < x^-_{1_L} < x^-_{1_R}, \tag{5.4.9}$$

which in regions (III), (IV), and (V), that is for $t < \frac{1}{2}(b-a)$, they are instead ordered as

$$x^+_{1_L} < x^+_{2_L} < x^+_{1_R} < x^+_{2_R}, \qquad x^-_{2_L} < x^-_{1_L} < x^-_{2_R} < x^-_{1_R}. \tag{5.4.10}$$

Using this, it is straightforward to determine the generalised entropy for the no island contribution in the high temperature limit. The exact expression for the entropy is [79]

$$S_{\text{QFT}}(A_L \cup A_R) = -\frac{c}{6}\sum_{\mu<\nu}(-1)^{\mu-\nu}\log\left[-(w_\mu^+ - w_\nu^+)(w_\mu^- - w_\nu^-)\right] - \frac{c}{12}\sum_\mu \log|w_\mu^+ w_\mu^-|. \tag{5.4.11}$$

Here $\mu, \nu = 1, 2, 3, 4$ correspond to the endpoints of the intervals, labelled $2_L, 1_L, 1_R, 2_R$. The last term on the RHS is the contribution from the Weyl anomaly which comes from rescaling the flat metric in the reservoir to the flat metric in the Kruskal frame, that is $-\mathrm{d}x^+\mathrm{d}x^- \to -\mathrm{d}w^+\mathrm{d}w^-$. In the high temperature limit it's useful to observe that if $x_\mu^\pm > x_\nu^\pm$, then

$$\log(w_\mu^+ - w_\nu^+) \approx \log w_\mu^+ \approx \frac{2\pi}{\beta}x_\mu^+, \qquad \log(w_\nu^- - w_\mu^-) \approx \log w_\nu^- \approx -\frac{2\pi}{\beta}x_\nu^-, \tag{5.4.12}$$

up to corrections which are exponentially suppressed in the high temperature limit. More precisely, what is meant here by the high temperature limit is the limit where the lengths of the intervals and their distance from the interface is taken to be much larger than the thermal scale. In addition, we work in the limit where the effective gravitational coupling is weak, i.e. $s \ll 1$. As this corresponds to moving away from extremality, it is not clear that one retains the connection between JT gravity and near-extremal black holes in higher dimensions in this limit. Using this, and the ordering (5.4.9), we find that in regions (I) and (II) the generalised entropy is given by

$$S_{\text{gen}}(A_L \cup A_R) = \underbrace{\frac{\pi c}{3\beta}\left[(t+b)+(t+a)\right]}_{\text{left}} + \underbrace{\frac{\pi c}{3\beta}\left[(t+b)+(t+a)\right]}_{\text{right}} - \underbrace{\frac{\pi c}{3\beta}(2a+2b)}_{\text{Weyl factor}} = \frac{4\pi c t}{3\beta}. \tag{5.4.13}$$

Notice that the left-moving and right-moving contributions to the entropy are equal since we focus on the symmetric case where A_L is an identical copy of A_R. With the ordering (5.4.10), we find that in regions (III), (IV), and (V) the generalised entropy is given by

$$S_{\text{gen}}(A_L \cup A_R) = \frac{2\pi c}{3\beta}\left[(t+b)+(-t+b)\right] - \frac{\pi c}{3\beta}(2a+2b) = \frac{2\pi c(b-a)}{3\beta}. \tag{5.4.14}$$

As expected, the entropies (5.4.13) and (5.4.14) are equal when $t = \frac{1}{2}(b-a)$ which is the boundary between temporal regimes (II) and (III). Also, this matches precisely with the first two contribution to the entropy (5.3.25) in the BCFT model which arise from the connected channels (2) and (1) in figure 5.9.

The other two contributions to the entropy (5.3.25) in the BCFT model arise from partially disconnected channels which involved one-point functions for two out of the four twist and antitwist fields. This suggests that there should be an island I with two

QESs, that is, I is a single interval. The generalised entropy for such a configuration is

$$\begin{aligned} S_{\text{gen}}(I \cup A_L \cup A_R) = & \underbrace{2S_0 + \frac{2\pi\phi_r/\beta}{4G} \sum_{\mu=1_Q,2_Q} \frac{1 - w_\mu^+ w_\mu^-}{1 + w_\mu^+ w_\mu^-}}_{\text{dilaton}} \\ & - \frac{c}{6} \sum_{\mu<\nu} (-1)^{\mu-\nu} \log[-(w_\mu^+ - w_\nu^+)(w_\mu^- - w_\nu^-)] \\ & - \underbrace{\frac{c}{12} \sum_{\mu=2_L,1_L,1_R,2_R} \log|w_\mu^+ w_\mu^-|}_{\text{reservoir Weyl factor}} - \underbrace{\frac{c}{6} \sum_{\mu=1_Q,2_Q} \log(1 + w_\mu^+ w_\mu^-)}_{\text{AdS}_2 \text{ Weyl factor}} . \end{aligned} \tag{5.4.15}$$

Here $\mu, \nu = 1, 2, 3, 4, 5, 6$ correspond to the endpoints of the intervals, labelled 2_L, 1_L, 1_Q, 2_Q, 1_R, 2_R which now involve two QESs. Now there is also a contribution from the Weyl anomaly which comes from rescaling the AdS metric to the flat metric in the Kruskal frame.

In general solving for the extrema of the generalised entropy would only be possible numerically. However, there are two key simplifications which emerge in the high temperature limit. Firstly, the QESs remain close to the horizon in the semiclassical limit, that is $|w_\sigma^+ w_\sigma^-|$ for $\sigma = 1_Q, 2_Q$. This allows us to simplify the dilaton term and write the equations for the QESs as

$$w_\sigma^\mp = -s \sum_{\mu \neq \sigma} \frac{(-1)^{\sigma-\mu}}{w_\sigma^\pm - w_\mu^\pm}, \tag{5.4.16}$$

where $s \ll 1$ was defined in (5.4.5). The second simplification is that in the high temperature limit the behaviour of the solution can separated into three temporal regimes (II), (III), and (IV), mirroring the BCFT calculation. In each of these regimes only one of the terms on the RHS of (5.4.16) dominate. Subleading terms are responsible for determining whether the QESs lie just behind or just outside the horizon.

Regime (IIa): We first consider the case that $2a < (b-a)/2$, so that t_{exit}, defined in (5.4.6), lies in temporal regime (II). Then regime (II) can be split into two subregimes:

$$\text{(IIa)} = \{a < t \leq 2a\}, \qquad \text{(IIb)} = \{2a < t < \frac{1}{2}(b-a)\}. \tag{5.4.17}$$

The early time period (IIa) turns out to be captured by the same QESs locations (5.4.4) as in the problem where A_L and A_R were half lines. In this regime, keeping only the dominant terms, and assuming $|w_{1_Q}^-| \gg 1$, $|w_{1_Q}^+ w_{1_Q}^-| \ll 1$, the equation for the QES (5.4.16) reads

$$w_{1_Q}^- \approx \frac{s}{w_{1_Q}^+ - w_{1_L}^+}, \qquad \frac{1}{s} w_{1_Q}^+ \approx -\frac{1}{w_{1_L}^-} + \frac{1}{w_{1_Q}^-} - \frac{1}{w_{1_Q}^- - w_{1_R}^-}. \tag{5.4.18}$$

Because of the symmetry of the problem $w^{\pm}_{1_Q} = w^{\mp}_{2_Q}$. Explicitly, the solution for the QES is

$$w^{+}_{1_Q} = -se^{-2\pi(t+a)/\beta} + \frac{1}{s}e^{-6\pi(t-a)/\beta}, \qquad w^{-}_{1_Q} = se^{2\pi(t-a)/\beta}. \tag{5.4.19}$$

In each of the remaining temporal regimes, the solutions are dominated by only one term on the right-hand side of (5.4.16), say $\mu = \mu(\sigma)$,

$$w^{\mp}_{\sigma} \approx \frac{s}{w^{\pm}_{\sigma} - w^{\pm}_{\mu(\sigma)}}. \tag{5.4.20}$$

The terms that have been dropped are either subleading in the high temperature limit or cancel out in pairs. The simpler equations (5.4.20) are quadratic, with only one of the solutions consistent with the ansatz that the QESs lie close to the horizon,

$$w^{\pm}_{\sigma} = \frac{w^{\pm}_{\mu(\sigma)}}{2}\left(1 - \sqrt{1 + \frac{4s}{w^{+}_{\mu(\sigma)}w^{-}_{\mu(\sigma)}}}\right). \tag{5.4.21}$$

In the high temperature limit this solutions has a sharp crossover between two distinct regimes. Firstly, when $x^{+}_{\mu(\sigma)} > x^{-}_{\mu(\sigma)}$, the solution becomes

$$w^{\pm}_{\sigma} \approx -\frac{s}{w^{\mp}_{\mu(\sigma)}}. \tag{5.4.22}$$

According to (5.4.22), in this regime, a QES corresponds to the following pattern of coordinates:

$$\begin{array}{ccccc} \cdots < x^{+}_{\sigma} & < & x^{+}_{\mu(\sigma)} & < \cdots \\ \| & & \| & \\ \cdots < x^{-}_{\mu(\sigma)} & < & x^{-}_{\sigma} & < \cdots \end{array} \tag{5.4.23}$$

For these solutions it follows from (5.4.22) that $\pm x^{\pm}_{\sigma}$ takes values in the set $\{-t-b, -t-a, t-b, t-a\}$. The second regime is where $x^{+}_{\mu(\sigma)} < x^{-}_{\mu(\sigma)}$, for which the solution becomes

$$w^{\pm}_{\sigma} \approx \sqrt{s\frac{w^{\pm}_{\mu(\sigma)}}{w^{\mp}_{\mu(\sigma)}}}\,. \tag{5.4.24}$$

In this case, the pattern is:

$$\begin{array}{c} \cdots < x^{+}_{\mu(\sigma)} < x^{+}_{\sigma} < \cdots \\ \| \\ \cdots < x^{-}_{\sigma} < x^{-}_{\mu(\sigma)} < \cdots \\ \| \\ \frac{1}{2}\left(x^{+}_{\mu(\sigma)} + x^{-}_{\mu(\sigma)}\right) \end{array} \tag{5.4.25}$$

Regime (IIb): In this regime the solution is of the type (5.4.22) with

$$w^{\pm}_{1_Q} = -\frac{s}{w^{\mp}_{1_L}} = \mp s e^{2\pi(\mp t-a)/\beta}, \qquad w^{\pm}_{2_Q} = -\frac{s}{w^{\mp}_{1_R}} = \pm s e^{2\pi(\pm t-a)/\beta}. \tag{5.4.26}$$

The QESs lie outside the horizon as show in figure 5.13 and the orderings of the $x^{\pm}$ coordinates are

$$x^{+}_{1_Q} < x^{+}_{1_L} < x^{+}_{2_Q} < x^{+}_{1_R} < x^{+}_{2_L} < x^{+}_{2_R}, \qquad x^{-}_{2_L} < x^{-}_{2_R} < x^{-}_{1_L} < x^{-}_{1_Q} < x^{-}_{1_R} < x^{-}_{2_Q}, \tag{5.4.27}$$

where the coordinates of the left and right QESs are shown in red and blue, respectively. One can also check that solution for the QESs is consistent with the ansatz that they lie close to the horizon since

$$|w^{+}_{1_Q} w^{-}_{1_Q}| = |w^{+}_{2_Q} w^{-}_{2_Q}| = s^2 e^{-4\pi a/\beta} \ll 1, \tag{5.4.28}$$

and the QES come in the pattern (5.4.23).

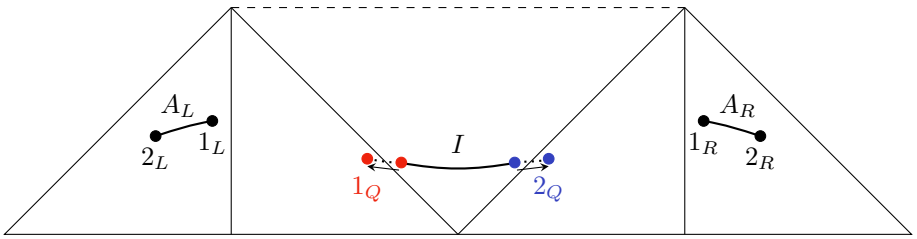


Figure 5.13: In region (II) the QESs, shown in red and blue, move from inside the horizon to outside if $2a < \frac{1}{2}(b-a)$. If $2a > \frac{1}{2}(b-a)$, they remain inside the horizon.

Evaluating the generalised entropy in temporal region (II) we find

$$\begin{aligned} S_{\text{gen}}(I \cup A_L \cup A_R) &= 2S_{\text{BH}} + \frac{2\pi c}{3\beta}\left[(t+b) + (t+a) + (-t+a)\right] - \frac{\pi c}{3\beta}(2a+2b) \\ &= 2S_{\text{BH}} + \frac{2\pi c}{3\beta}(t+a). \end{aligned} \tag{5.4.29}$$

Since the QESs lie close to the horizon, each contribution from the dilaton is simply the Bekenstein-Hawking entropy. This precisely matches with the contribution (5.3.25) in the BCFT model which comes from the partially disconnected channel (3) in figure 5.9. Even though the functional form of the solution for the QESs in regimes (IIa) and (IIb) were different, up to exponentially small corrections, they lead to the same contribution for the generalised entropy.

This suggests an identification between the boundary entropy $\log g_b$ and the Bekenstein-Hawking entropy S_{BH}:

$$\log g_b \quad \Leftrightarrow \quad S_{\text{BH}}. \tag{5.4.30}$$

The Bekenstein-Hawking entropy can be viewed as a sort of temperature dependent boundary entropy. Moreover, the free fermion theory which has vanishing boundary entropy can be thought of as an infrared limit in which typical length and time scales, measured in units of the inverse temperature, are taken to be much larger than the Bekenstein-Hawking entropy.

When there is an island, the QFT entropy term which contributes to the generalised entropy (5.4.15) is determined by a six-point function of twist and antitwist fields in the bulk matter CFT. Given a configuration of QESs we can then ask what the dominant channel for the OPE is. Within region (II) we can identify how the dominant *bulk* OPE channel is related to the dominant OPE channel in the BCFT model. Even though the function form for the solution in region (II) changes over time, the entropy is always determined by the OPE limit

$$w^{\pm}_{1_L} \to w^{\pm}_{1_Q}, \qquad w^{\pm}_{1_R} \to w^{\pm}_{2_Q}, \qquad w^{\pm}_{2_L} \to w^{\pm}_{2_R}. \tag{5.4.31}$$

In this limit the correlator of twist and antitwist fields factorise into a product of two-point functions which leads to the contribution

$$S_{\text{gen}}(I \cup A_L \cup A_R) \supset \frac{c}{6} \log |w^{+}_{1_L,1_Q} w^{-}_{1_L,1_Q} w^{+}_{2_Q,1_R} w^{-}_{2_Q,1_R} w^{+}_{2_L,2_R} w^{-}_{2_L,2_R}|, \tag{5.4.32}$$

where $w^{\pm}_{\mu\nu} = w^{\pm}_{\mu} - w^{\pm}_{\nu}$. Including the contribution from the dilaton and the Weyl anomaly gives (5.4.29). Notice how a contraction between a twist (or antitwist) field in the reservoir and antitwist (or twist) field in the AdS_2 region reproduces the contraction between a point and its image in the BCFT model (see panel (3) of figure 5.9).

Regime (III): The solution in region (III) can be found by understanding how the ordering (5.4.27) changes as we move into region (III) and then using this new ordering as an ansatz to solve (5.4.16) for the QESs. We find that the solution is of type (5.4.22) with

$$\begin{aligned} w^{-}_{1_Q} &= -\frac{s}{w^{+}_{1_L}} = s e^{2\pi(t-a)/\beta}, & w^{+}_{1_Q} &= -\frac{s}{w^{-}_{2_R}} = s e^{2\pi(t-b)/\beta}, \\ w^{-}_{2_Q} &= -\frac{s}{w^{+}_{2_L}} = s e^{2\pi(t-b)/\beta}, & w^{+}_{2_Q} &= -\frac{s}{w^{-}_{1_R}} = s e^{2\pi(t-a)/\beta}. \end{aligned} \tag{5.4.33}$$

We see that as we move from region (II) to (III) the QESs move from outside the horizon to inside the horizon, as shown in figure 5.14, if $2a < \frac{1}{2}(b-a)$, or simply remain inside if $2a < \frac{1}{2}(b-a)$. Later we will see more precisely how this happens.

The ordering of the $x^{\pm}$ coordinates in region (III) are

$$x^{+}_{1_Q} < x^{+}_{1_L} < x^{+}_{2_Q} < x^{+}_{2_L} < x^{+}_{1_R} < x^{+}_{2_R}, \qquad x^{-}_{2_L} < x^{-}_{1_L} < x^{-}_{2_R} < x^{-}_{1_Q} < x^{-}_{1_R} < x^{-}_{2_Q}, \tag{5.4.34}$$

and the QESs remain close to the horizon,

$$|w^{+}_{1_Q} w^{-}_{1_Q}| = |w^{+}_{2_Q} w^{-}_{2_Q}| = s^2 e^{2\pi(2t-a-b)/\beta} \ll 1, \tag{5.4.35}$$

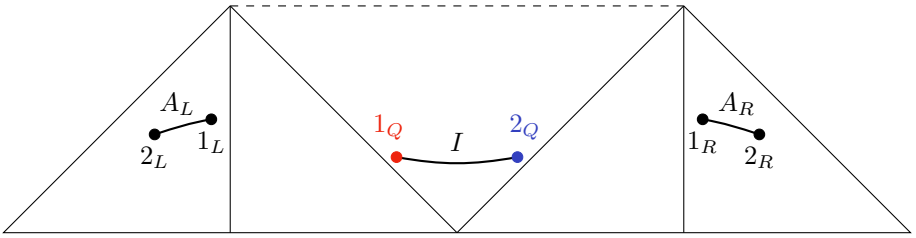


Figure 5.14: In regions (III) and (IV) the QESs, shown in red and blue, lie inside the horizon.

since $t < \frac{1}{2}(a+b)$ in region (III). The entropy in regime (III) is

$$\begin{aligned} S_{\text{gen}}(I \cup A_L \cup A_R) &= 2S_{\text{BH}} + \frac{2\pi c}{3\beta}\left[(t+b)+(-t+b)+(-t+a)\right] - \frac{\pi c}{3\beta}(2a+2b) \\ &= 2S_{\text{BH}} + \frac{2\pi c}{3\beta}(-t+b). \end{aligned} \tag{5.4.36}$$

With the identification between the boundary entropy and Bekenstein-Hawking entropy, this matches precisely with the BCFT result (5.3.25). In the BCFT model this contribution arises from a contraction between a twist (and antitwist) field in the left copy and an antitwist (and twist) field in the right copy. As mentioned earlier, these are lightcone singularities. In the bulk, this contribution arises from the limits

$$w^+_{1_L} \to w^+_{1_Q}, \quad w^-_{1_Q} \to w^-_{2_R}, \quad w^+_{2_L} \to w^+_{2_Q}, \quad w^-_{2_Q} \to w^-_{1_R}, \quad w^+_{1_R} \to w^+_{2_R}, \quad w^-_{2_L} \to w^-_{1_L}. \tag{5.4.37}$$

Notice that the $w^\pm$ coordinates are in different channels, so this does not correspond to an OPE limit. Instead, it corresponds to a lightcone singularity in the bulk. The leads to the contribution

$$S_{\text{gen}}(I \cup A_L \cup A_R) \supset \frac{c}{6}\log|w^+_{1_L,1_Q} w^-_{1_Q,2_R} w^+_{2_L,2_Q} w^-_{2_Q,1_R} w^+_{1_R,2_R} w^-_{2_L,1_L}|. \tag{5.4.38}$$

Including the contribution from the dilaton and the Weyl anomaly gives (5.4.29).

Regime (IV): Finally, the solution for the QESs in region (III) transitions into region (IV) as a solution of type (5.4.24),

$$w^\pm_{1_Q} = \sqrt{s}e^{\mp\pi(b-a)/\beta}, \qquad w^\pm_{2_Q} = \sqrt{s}e^{\pm\pi(b-a)/\beta}. \tag{5.4.39}$$

The QESs remain inside the horizon as shown in figure 5.14. Again, the QESs remain close to the horizon since

$$|w^+_{1_Q} w^-_{1_Q}| = |w^+_{2_Q} w^-_{2_Q}| = s \ll 1, \tag{5.4.40}$$

The ordering of the $x^\pm$ coordinates are

$$x^+_{1_L} < x^+_{1_Q} < x^+_{2_L} < x^+_{2_Q} < x^+_{1_R} < x^+_{2_R}, \quad x^-_{2_L}, < x^-_{1_L} < x^-_{1_Q} < x^-_{2_R} < x^-_{2_Q} < x^-_{1_R}, \tag{5.4.41}$$

and the come in the pattern (5.4.25). The entropy in regime (IV) is

$$\begin{aligned} S_{\text{gen}}(I \cup A_L \cup A_R) &= 2S_{\text{BH}} + \frac{2\pi c}{3\beta}\left[(t+b) + \tfrac{1}{2}(b-a) + \tfrac{1}{2}(a-b)\right] - \frac{\pi c}{6\beta}(4a+4b) \\ &= 2S_{\text{BH}} + \frac{2\pi c}{3\beta}(t-a), \end{aligned} \tag{5.4.42}$$

which matches with the rising portion of the contribution from channel (4) in figure 5.3.25 in the BCFT model. As in the BCFT model, both the dip and rise of regions (III) and (IV) are governed by the same limits 5.4.37, despite the fact that the functional form of the solutions for the QESs change.

The entropy $S(A_L \cup A_R)$ is then given by minimising over the no island and island contributions, which reproduces the BCFT result 5.3.25. The final curve is shown in figure 5.3, with the no island contribution shown in black and the island contribution shown in red, for the case that the Bekenstein-Hawking entropy is negligible.

Regions (IIa) and (IIb) show no qualitative difference insofar as the entropy is concerned. When $2a > \frac{1}{2}(b-a)$, the two regions merge into one and the QESs remain inside the horizon at all times. Hence, in this case regime (IIb) is eliminated and the analysis of regions (III) and (IV) remains largely unchanged, except for the transient position of the QES in regime (III) which receives a correction, whilst remaining inside the horizon.

What our approximation (5.4.20) to the equations for the QESs (5.4.16) obscures is that the island contributions in regions (II), (III) and (IV) are in fact smoothly related across the different regions in a way that becomes sharper as the temperature increases. We can do better by working with a refinement of the approximation (5.4.20) which includes all the relevant terms for describing the island across regions (II), (III) and (IV):

$$\begin{aligned} w^-_{1_Q} &= \frac{s}{w^+_{1_Q} - w^+_{1_L}}, & w^+_{1_Q} &= \frac{s}{w^-_{1_Q}} - \frac{s}{w^-_{1_Q} - w^-_{1_R}} + \frac{s}{w^-_{1_Q} - w^-_{2_R}} - \frac{s}{w^-_{1_L}}, \\ w^+_{2_Q} &= \frac{s}{w^-_{2_Q} - w^-_{1_R}}, & w^-_{2_Q} &= \frac{s}{w^+_{2_Q}} - \frac{s}{w^+_{2_Q} - w^+_{1_L}} + \frac{s}{w^+_{2_Q} - w^+_{2_L}} - \frac{s}{w^+_{1_R}}. \end{aligned} \tag{5.4.43}$$

In the last terms, we have used the fact that $|w^-_{1_Q}| \ll |w^-_{1_L}|$, $|w^+_{2_Q}| \ll |w^+_{1_R}|$, $|w^+_{1_Q}| \ll |w^-_{1_Q}|$, and $|w^-_{2_Q}| \ll |w^+_{2_Q}|$, respectively. Remarkably, each pair of equations simply results in a quadratic equation for w^+_{1Q} and w^-_{2Q}:

$$\begin{aligned} \frac{1}{w^-_{1_Q} - w^-_{2_R}} - \frac{1}{w^-_{1_Q} - w^-_{1_R}} - \left(\frac{1}{w^-_{1_L}} + \frac{1}{s} w^+_{1_L}\right) &= 0, \\ \frac{1}{w^+_{2_Q} - w^+_{2_L}} - \frac{1}{w^+_{2_Q} - w^+_{1_L}} - \left(\frac{1}{w^+_{1_R}} + \frac{1}{s} w^-_{1_R}\right) &= 0. \end{aligned} \tag{5.4.44}$$

The solutions to these equations smoothly interpolate between the asymptotic solutions in regions (II), (III), and (IV). For example, at high temperatures, the solution for the w^- coordinate of the right QES can be written as

$$w^-_{2_Q} = 2s\left(\sqrt{4s^2e^{2\pi(b-3a)/\beta} + 4se^{2\pi(b-a)/\beta} + e^{4\pi(b-t)/\beta} - 2e^{2\pi(b+a-2t)/\beta}} - e^{2\pi(a-t)/\beta} - e^{2\pi(b-t)/\beta}\right)^{-1} - e^{-2\pi(t-a)/\beta}. \tag{5.4.45}$$

The solutions is valid through the regions (II)→(III)→(IV) with transitions between the different asymptotic solutions at $t = \frac{1}{2}(b \pm a)$ as t increases. In particular, for the case that $2a < \frac{1}{2}(b-a)$, the QES moves out from inside the horizon at $t = 2a$ and as region (II) transitions into region (III) at $t = \frac{1}{2}(b-a)$, the QES moves smoothly from outside to inside the horizon, as shown in figure 5.15. The left QES behaves in a similar way.

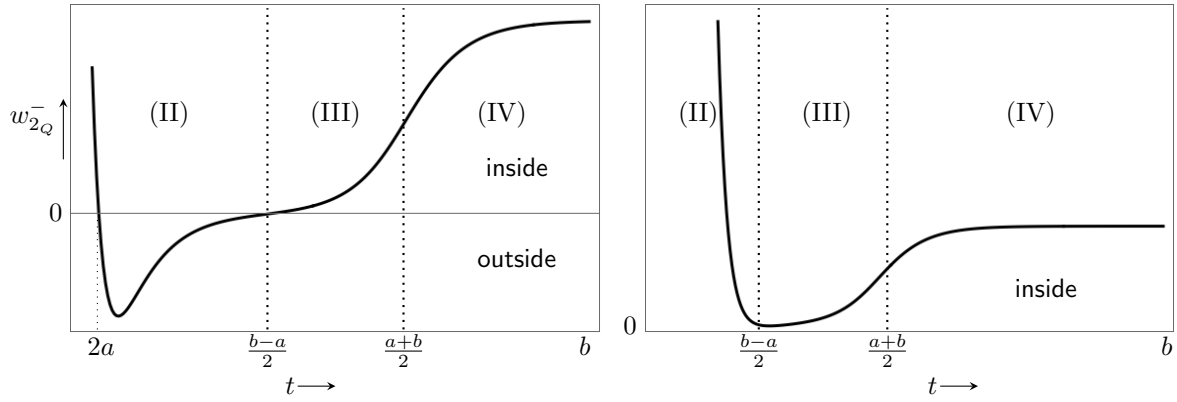


Figure 5.15: The solution for the w^- coordinate of the right QES determined numerically for large but finite temperature. When $2a < \frac{1}{2}(b-a)$ (left), the QES starts inside the horizon, exits at $t = 2a$ at the onset of region (II), and plunges back in at $t = \frac{1}{2}(b-a)$ and remains inside through regions (III) and (IV). For $2a > \frac{1}{2}(b-a)$ (right) the QES remains inside the horizon at all times.

5.4.3 Small intervals

For $b < 3a$ there is a different pattern for the behaviour of the entropy

$$\begin{aligned} \text{(I)} &= \{0 < t < \tfrac{1}{2}(b-a)\}, & \text{(V)} &= \{\tfrac{1}{2}(b-a) < t < a\}, \\ \text{(III)} &= \{a < t < \tfrac{1}{2}(a+b)\}, & \text{(IV)} &= \{\tfrac{1}{2}(a+b) < t < b\}, \\ \text{(V)} &= \{b < t\}. \end{aligned} \tag{5.4.46}$$

Here the labelling matches the labelling of the temporal regions in the previous section. In this case, the entropy plateaus early at $t = \frac{1}{2}(b-a)$ and dips later on before returning to a plateau. For this case, the QESs remain inside the horizon since region (II) does not occur. The entropy is shown in figure 5.10.

5.4.4 Generic case

For the generic case with two intervals A_L and A_R, that is, A_L is not an identical copy of A_R, we have seen that there are additional channels which can contribute and give rise to the facet-like structure of (5.12). In particular, the crossover between regions (II) and (III) gets replaced by a new region which is governed by a new disconnected BCFT channel, shown in panel (5) of 5.11. Based on our observations in the symmetric case, we expect channel (5) to correspond to island with two QESs with the QFT entropy term governed by the bulk limits:

$$w^+_{1_L} \to w^+_{1_Q}, \quad w^-_{1_Q} \to w^-_{2_R}, \quad w^+_{1_R} \to w^+_{2_Q}, \quad w^-_{2_Q} \to w^-_{1_R}, \quad w^+_{2_L} \to w^+_{2_R}, \quad w^-_{2_L} \to w^-_{1_L}. \tag{5.4.47}$$

A quick calculation confirms that this is indeed the case. The equations for the QESs are still (5.4.43) and the solution for the w^- coordinate of the right QES can be obtained in an analogous to (5.4.45).

5.5 Geodesic approximation

In the high temperature limit the entanglement structure can be visualised in terms of the positions of localised wave packets, as modes with characteristic energy $1/\beta$ can be localised over distance scales of order β, which is small for high temperatures. A left-moving wave packet on the left, localised around the geodesic, or ray, $x^+ = \lambda$ is entangled with a localised wave packet on the right with the reflected coordinates $x^+ = \lambda$. There is a similar entanglement structure for the right-moving modes. In this section we can analyse the entropy of $A_L \cup A_R$ by a simple method of ray tracing.

5.5.1 BCFT

Consider left-moving modes in the BCFT that pass through A_R at time t (see figure 5.16). There are two distinct sets of modes: the left-moving modes in the interval $x^+ \in A^{(l)}_R \coloneqq [a+t, b+t]$ that pass through A_R as left-moving modes, and the left-moving modes in interval $x^+ \in \hat{A}^{(l)}_R \coloneqq [\max(t-b,0), \max(t-a,0)]$ that reflect off the boundary and pass through A_R as right-moving modes. On the left, there are no reflected modes to consider and the left-moving modes that pass through A_L at time t are $x^+ \in A^{(l)}_L \coloneqq [\max(a-t,0), \max(b-t,0)]$. The three subsets of modes are shown in figure 5.16. There is a similar story for the right-moving modes.

The modes on the left are entangled with modes on the right under the reflection $x^+ \to -x^+$, which we define as

$$x^+ \in \tilde{A}^{(l)}_L = [\max(a-t,0), \max(b-t,0)]. \tag{5.5.1}$$

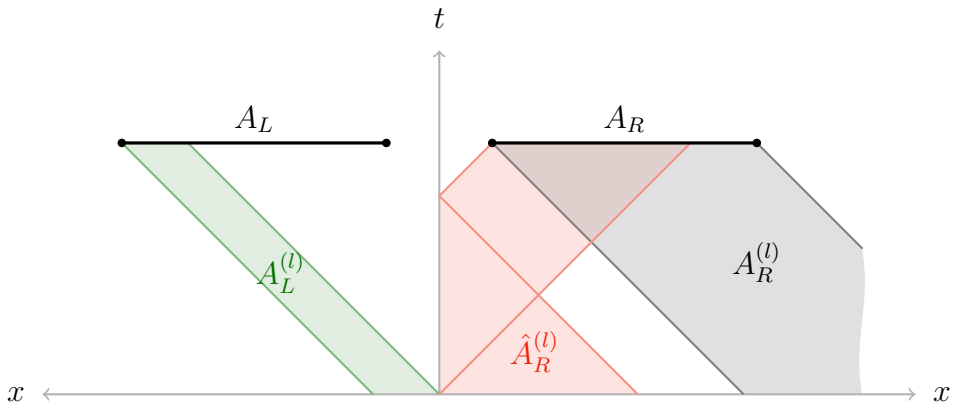


Figure 5.16: The left-moving modes that pass through $A_L \cup A_R$ at time t are split into the three subsets as indicated. Here, t is in the temporal region (IV).

The entropy of $A_L \cup A_R$ is then

$$S_{\text{BCFT}}(A_L \cup A_R) = S_{\text{rad}}\left(\tilde{A}_L^{(l)} \ominus (A_R^{(l)} \cup \hat{A}_R^{(l)})\right) + S_{\text{rad}}\left(\tilde{A}_R^{(r)} \ominus (A_L^{(r)} \cup \hat{A}_L^{(r)})\right), \quad (5.5.2)$$

where we have added a similar contribution from the right-moving modes. The entropy of a null interval of radiation is defined in (5.3.17). The entropy can be read off from 5.17 which illustrates the effect of the symmetric difference.

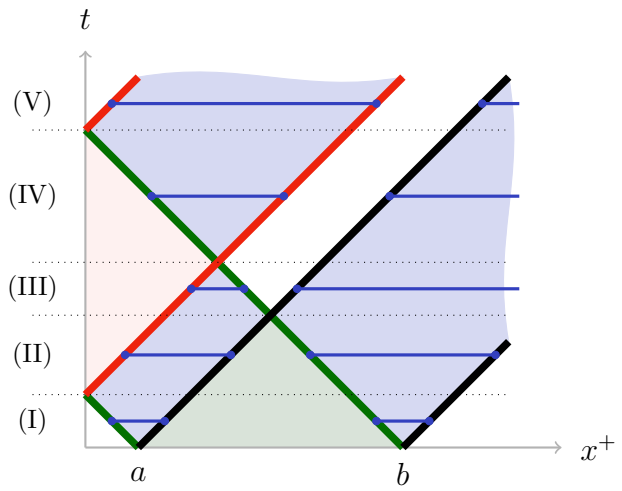


Figure 5.17: The different sets of modes: $A_R^{(l)}$ bordered in black, the reflected modes $\hat{A}_R^{(l)}$ in red, and $\tilde{A}_L^{(l)}$ in green. The entropy is determined by the symmetric difference $A_R^{(l)} \ominus \tilde{A}_L^{(l)}$ (blue).

5.5.2 JT gravity

In this section we use a similar geometrical approach to visualise the entanglement structure and calculate the entropy in the JT model. This visualisation of the entanglement structure is related to the island in the stream recipe [57, 58], discussed in chapter 3.

We denote the projection of the right-moving modes that pass through A_R onto right null infinity $\mathcal{I}_R^+$ as $A_R^{(r)}$, as shown in figure 5.18. In the high temperature limit, the entropy of these modes is given by the thermodynamic formula (5.3.17). However, the interval $A_L \cup I$ contains some right-moving which are entangled with the modes in $A_R^{(r)}$ and so the contribution of these modes to the entropy effectively cancels out. The entangled modes are related by the reflection $w^- \to -w^-$. Denoting the reflected right moving modes $A^{(r)}$ as $\tilde{A}^{(r)}$, we can map the right-moving modes that pass through $A_L \cup I$ onto modes $\tilde{A}_L^{(r)} \cup \tilde{I}^{(r)}$ on right null infinity $\mathcal{I}^+$ by a combined reflection and projection. Note that the reflection preserves the values of the coordinates x^- and so it is simple to find the images of I and A_L on $\mathcal{I}_R^+$, they are simply given by the labelled by the locations of their x^- coordinates.

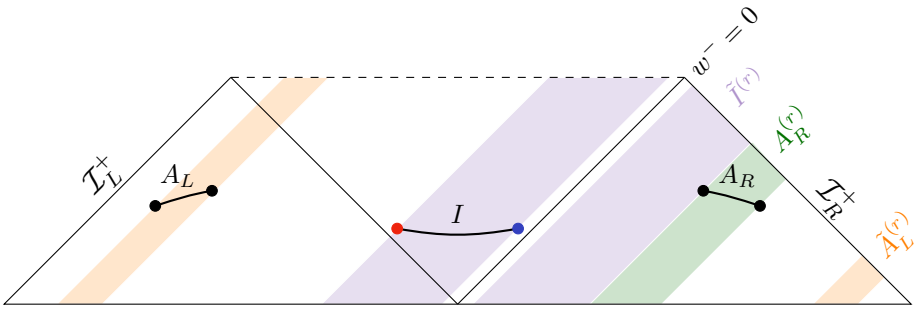


Figure 5.18: The right-moving modes that pass through A_R can be projected onto right null infinity $\mathcal{I}_R^+$. The right-moving modes that pass through $A_L \cup I$ can be mapped to $\mathcal{I}_R^+$ by a combined reflection and projection.

There is a slight subtlety for the islands. If a QES is inside the horizon with $w_\sigma^- = e^{-2\pi x_\sigma^-/\beta}$, the reflection $w^- \to -w^-$ maps it to $-e^{-2\pi x_\sigma^-/\beta}$ and projects onto $\mathcal{I}_R^+$ as a point with null coordinate x_σ^-. However, if a QES is outside the horizon with $w_\sigma^- = -e^{-2\pi x_\sigma^-/\beta}$, then the state of the modes in null interval $[-w_\sigma^-, w_\sigma^-]$ is pure and so doesn't contribute to the entropy. So, as far as far as the entropy is concerned, we can replace the coordinate of the QES by $e^{-2\pi x_\sigma^-/\beta}$, which lies inside the horizon.

The net contribution from the right-moving modes is determined by the thermodynamic formula (5.3.17) for the interval $A_R^{(r)} \ominus (\tilde{A}_L^{(r)} \cup \tilde{I}^{(r)})$. The symmetric difference accounts for the right-moving modes in $A_L \cup I$ which purify the right-moving modes in A_R. Applying the same reasoning to the right-moving modes leads to the generalised

entropy

$$S_{\text{gen}}(A_L \cup A_R) = 2S_{\text{BH}} + S_{\text{rad}}\left(A_R^{(r)} \ominus (\tilde{A}_L^{(r)} \cup \tilde{I}^{(r)})\right) + S_{\text{rad}}\left(A_L^{(l)} \ominus (\tilde{A}_R^{(l)} \cup \tilde{I}^{(l)})\right) + S_{\text{Weyl}}, \tag{5.5.3}$$

where

$$S_{\text{Weyl}} = \frac{c}{12} \sum_{\sigma=1_Q,2_Q} \log |w_\sigma^+ w_\sigma^-|, \tag{5.5.4}$$

accounts for the fact that the QFT entropy is not purely given by the thermodynamic formula as the geometry is not flat everywhere due to curvature of the AdS_2 region. The set of modes $A_R^{(r)} \ominus (\tilde{A}_L^{(r)} \cup \tilde{I}^{(r)})$ is illustrated in figure 5.19. For example, consider region (III), where the entropy dips. Here the projections on $\mathcal{I}_R^+$ in terms of the x^- coordinates are

$$A_R^{(r)} = [t-b, t-a], \qquad \tilde{A}_L^{(r)} = [-t-b, -t-a], \qquad \tilde{I}_R^{(r)} = [-t+a, -t+b]. \tag{5.5.5}$$

Hence,

$$A_R^{(r)} \ominus (\tilde{A}_L^{(r)} \cup \tilde{I}^{(r)}) = [-t-b, -t-a] \cup [t-b, -t+a] \cup [t-a, -t+b] \tag{5.5.6}$$

and so the entropy is

$$S_{\text{rad}}(A_R^{(r)} \ominus (\tilde{A}_L^{(r)} \cup \tilde{I}^{(r)})) = \frac{\pi c}{6\beta}\left[(b-a) + (b+a-2t) + (b+a-2t)\right]. \tag{5.5.7}$$

The left-moving modes give an identical contribution. Note that the size of the second and third intervals decreases as time increases, which corresponds to the dip in the entropy. Here S_{Weyl} is given by

$$S_{\text{Weyl}}(1_Q) = S_{\text{Weyl}}(1_Q) = \frac{\pi c}{6\beta}(2t-a-b). \tag{5.5.8}$$

Adding all the contributions gives the result (5.4.36).

5.6 Holographic BCFTs

In both the BCFT and JT model we have seen that the entropy experiences a dip at intermediate times. The mechanism for the dip can be understood as arising from a lightcone singularity in the correlator of twist and antiwtist fields which computes the entropy $S(A_L \cup A_R)$ in the BCFT model or the entropy $S_{\text{QFT}}(I \cup A_L \cup A_R)$ in the JT model. These singularities are not fixed by the OPE and so do not exist for all CFTs. In this section, we show that the dip, and hence the lightcone singularity, is absent in holographic BCFTs using Takayanagi's formula [109] for computing the entropy holographically. Although Takayanagi's proposal is valid in any dimension, we restrict to the case of interest, where the BCFT is two-dimensional and so the bulk is three-dimensional.

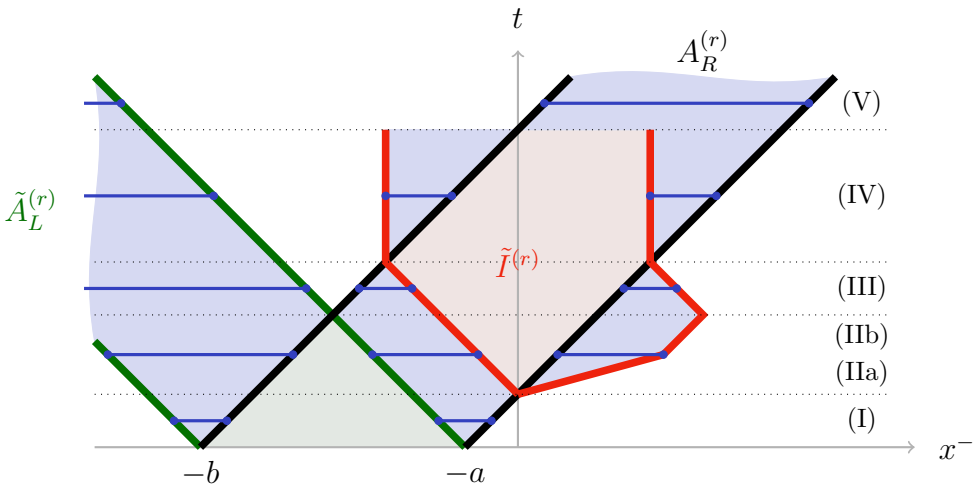


Figure 5.19: The different sets of modes: $A_R^{(r)}$ (black), $\tilde{A}_L^{(r)}$ (green), and $\tilde{I}^{(r)}$ (red). The latter two sets of modes purify the former, hence, the net contribution to the entropy is determined by the symmetric difference $A_R^{(r)} \ominus (\tilde{A}_L^{(r)} \cup \tilde{I}^{(r)})$ (blue). In region (III), the blue regions shrink due a collision between $\tilde{A}_L^{(r)}$ and $\tilde{I}^{(r)}$ as time increases. This accounts for the dip in the entropy.

5.6.1 Bulk geometry

For a holographic CFT on a manifold M with boundary ∂M, Takayanagi proposed that the dual bulk geometry should contain an *end of the world (EOW) brane* which ends on ∂M [109]. That is, in addition to the conformal boundary of AdS, the bulk geometry should contain an additional boundary, the EOW brane, where one imposes Neumann boundary conditions for the metric. In the simplest setting, where the action is taken to be the Einstein-Hilbert action together with the Gibbons-Hawking-York boundary term, there is now an additional contribution from the portion of the boundary term integrated along the EOW brane

$$\frac{1}{8\pi G}\int_{\text{EOW}} \mathrm{d}^2x\sqrt{-h}(K-T), \tag{5.6.1}$$

where h is the induced metric on the EOW brane and T is the tension of the brane. The Neumann boundary condition implies that $K = 2T$. From this equation, the location of the EOW can often be determined from symmetry alone. For example, for the vacuum state of a BCFT on the half space $x \geq 0$ the bulk geometry is the Poincaré patch of AdS_3 where the metric is

$$\mathrm{d}s^2 = \frac{-\mathrm{d}t^2 + \mathrm{d}x^2 + \mathrm{d}z^2}{z^2}. \tag{5.6.2}$$

We expect that the EOW brane can be parameterised as a curve in the (x, z)-plane times $\mathbb{R}$, which represents the time direction. The trace of the extrinsic curvature can

be written as

$$K = \frac{1}{2} h^{ab} \mathcal{L}_n h_{ab}, \tag{5.6.3}$$

where $\mathcal{L}$ is the Lie derivative along the outward pointing unit normal vector n to the brane. Since h^a_b is a projector onto tensors tangent to the brane, if $X^a = (0, \dot{x}, \dot{z})$ denotes the tangent to the curve and $Y^a = (1, 0, 0)$ denotes the tangent to the time direction, we can write h_{ab} as

$$h_{ab} = \frac{X_a X_b}{(X, X)} + \frac{Y_a Y_b}{(Y, Y)}. \tag{5.6.4}$$

Locally, the curve can be described by a graph $x = f(z)$. The form of f is fixed by scale transformations to be $f(z) = \alpha z$ for some $\alpha \in \mathbb{R}$. It's convenient to parameterise this constant as $\alpha = \tan\theta$, with $-\frac{\pi}{2} < \theta < \frac{\pi}{2}$, so the EOW brane lies along

$$x = z \tan\theta. \tag{5.6.5}$$

If the spacetime is the region $x - z\tan\theta \leq 0$, the outward pointing normal is

$$n = \frac{N}{(N, N)^{\frac{1}{2}}}, \qquad N = \frac{\partial}{\partial x} - \tan\theta \frac{\partial}{\partial z}. \tag{5.6.6}$$

We then find

$$K = 2\sin\theta, \tag{5.6.7}$$

so that $T = \sin\theta$.

5.6.2 Takayanagi's formula

Takayanagi's formula generalises Ryu and Takayanagi's formula which says that the entropy $S(A)$ of a subregion A in a CFT is given by the area of a minimal, or RT, surface homologous to A,

$$S(A) = \min_{\gamma} \frac{A(\gamma)}{4G}. \tag{5.6.8}$$

The main feature of Takayanagi's proposal is that for a BCFT, the minimal surface can end on the EOW brane. More specifically, when considering the homology constraint, one should view the brane as the bulk geometry, rather than an additional boundary.

The simplest example is to compute the entropy of an interval $A = [0, L]$ in BCFT on the half space $x \geq 0$. This is computed by a minimal surface with one endpoint on the conformal boundary at $x = L$ and one endpoint on the brane,

$$S(A) = \frac{1}{4G} \log \frac{L}{\epsilon} + \frac{1}{4G} \tanh^{-1} T. \tag{5.6.9}$$

Using the formula $G = 3/2c$ of Brown and Henneaux [120] this is precisely what we found in BCFT (5.3.12) if we identify

$$\log g_b = \frac{1}{4G} \tanh^{-1} T. \tag{5.6.10}$$

5.6.3 Thermofield double

We begin by finding the dual geometry and EOW brane for the case where we have two copies of a BCFT in the thermofield double state. Recall that correlators in the thermofield double can be obtained from thermal correlators by analytic continuation. In turn, thermal correlators can be obtained from vacuum correlators using conformal transformations. It's useful to think of the mapping between the UHP and half of the thermal cylinder in two steps: first we map from the UHP to the disk, and then we map from the disk to half of the thermal cylinder. We now apply this logic to obtain the EOW brane for the thermofield double, starting from the EOW brane (5.6.5) for the vacuum state of a BCFT on a half space.

Working in complex coordinates $w = \tau + \mathrm{i}x$ on the UHP, where τ is Euclidean time, a general fractional linear transformation

$$\varphi : \; w \mapsto \frac{aw+b}{cw+d}, \tag{5.6.11}$$

may be decomposed as $\varphi = \varphi_4 \circ \varphi_3 \circ \varphi_2 \circ \varphi_1$ where

$$\varphi_1(w) = w + \frac{d}{c}, \qquad \varphi_2(w) = -\frac{1}{w}, \qquad \varphi_3(w) = \frac{ad-bc}{c^2}w, \qquad \varphi_4(w) = w + \frac{a}{c}. \tag{5.6.12}$$

Each of these transformations φ_i extends to an (orientation preserving) isometry ψ_i of AdS_3

$$\begin{aligned} \psi_1(w,z) &= \left(w + \frac{d}{c}, z\right), & \psi_2(w,z) &= \left(-\frac{\bar{w}}{|w|^2+z^2}, \frac{z}{|w|^2+z^2}\right), \\ \psi_3(w,z) &= \left(\frac{ad-bc}{c^2}w, \left|\frac{ad-bc}{c^2}\right| z\right), & \psi_4(w,z) &= \left(w + \frac{a}{c}, z\right). \end{aligned} \tag{5.6.13}$$

The composition $\psi = \psi_4 \circ \psi_3 \circ \psi_2 \circ \psi_1$ is

$$\psi(w,z) = \left(\frac{a\left[w + \frac{\bar{c}}{\bar{d}}\left(|w^2| + z^2\right)\right] + b\left(1 + \frac{\bar{c}}{\bar{d}}\bar{w}\right)}{c\left[w + \frac{\bar{c}}{\bar{d}}\left(|w^2| + z^2\right)\right] + d\left(1 + \frac{\bar{c}}{\bar{d}}\bar{w}\right)}, \left|\frac{ad-bc}{c^2}\right| \frac{z}{\left|w + \frac{d}{c}\right|^2 + z^2}\right). \tag{5.6.14}$$

The transformation which maps the UHP $\mathrm{Im}(w) \geq 0$ to the unit disk is

$$w \mapsto \frac{w - \mathrm{i}}{w + \mathrm{i}}, \tag{5.6.15}$$

so the corresponding isometry is

$$w \mapsto \frac{w + \bar{w} + \mathrm{i}\left(|w|^2 + z^2 - 1\right)}{w - \bar{w} + \mathrm{i}\left(|w|^2 + z^2 + 1\right)}, \qquad z \mapsto \frac{2z}{|w + \mathrm{i}|^2 + z^2}. \tag{5.6.16}$$

This isometry maps the EOW brane $x = z\tan\theta$ to

$$|w|^2 + (z + \tan\theta)^2 = \sec^2\theta. \tag{5.6.17}$$

To summarise, for two copies of a BCFT in the thermofield double, the dual geometry has metric

$$\mathrm{d}s^2 = \frac{-\mathrm{d}t^2 + \mathrm{d}x^2 + \mathrm{d}z^2}{z^2}, \tag{5.6.18}$$

where we use the portion of the geometry with

$$-t^2 + x^2 + (z + \tan\theta)^2 \geq \sec^2\theta. \tag{5.6.19}$$

The boundary of this region is the EOW brane. The relation between the Poincaré coordinates and the physical coordinates (u_R, y_R) in the R BCFT is

$$t = e^{2\pi y_R/\beta}\sinh\frac{2\pi u_R}{\beta}, \qquad x = e^{2\pi y_R/\beta}\cosh\frac{2\pi u_R}{\beta}. \tag{5.6.20}$$

Similarly, for the coordinates (u_L, y_L) in the L BCFT we have

$$t = -e^{2\pi y_L/\beta}\sinh\frac{2\pi u_L}{\beta}, \qquad x = -e^{2\pi y_L/\beta}\cosh\frac{2\pi u_L}{\beta}. \tag{5.6.21}$$

To compare with the notation of the previous sections we should identify $x^{\pm}_{L,R} = u_{L,R} \pm y_{L,R}$.

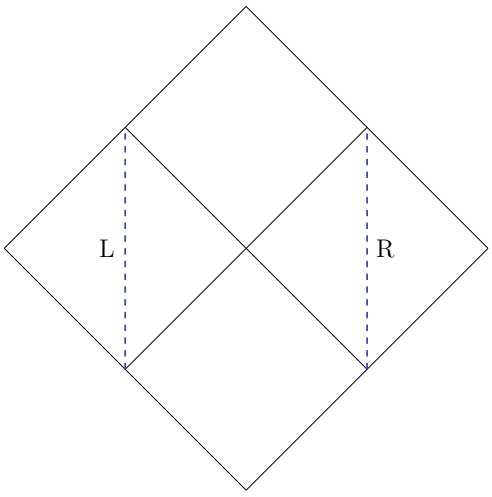


Figure 5.20: The Penrose diagram of the (t, x)-plane at $z = 0$. The blue dashed lines indicate where the EOW brane, which extends into the z-plane, intersects the boundary.

5.6.4 Endpoints on the boundary

First consider the case that the two endpoints p_1 and p_2 of the geodesic lie on the boundary. Let p_1 be the point $(t, x, z) = (t_1, x_1, 0)$ and p_2 be the point $(t, x, z) = (t_2, x_2, 0)$. There is an isometry which maps p_1 and p_2 to the points $(t, x, z) = (0, -r, 0)$ and $(t, x, z) = (0, r, 0)$. This isometry maps the geodesic to the semicircle centred at the origin

$$t = 0, \qquad z = \sqrt{r^2 - x^2}, \tag{5.6.22}$$

The length of the geodesic is

$$\ell(p_1, p_2) = 2 \tanh^{-1} \sqrt{1 - \frac{\epsilon_0^2}{r^2}} = 2 \log \frac{2r}{\epsilon_0} + O(\epsilon_0^2). \tag{5.6.23}$$

Since the map is an isometry $(2r)^2 = -(t_1 - t_2)^2 + (x_1 - x_2)^2$. So, we have

$$\ell(p_1, p_2) = 2 \log \frac{\sqrt{-(t_1 - t_2)^2 + (x_1 - x_2)^2}}{\epsilon_0} + \mathcal{O}(\epsilon_0^2). \tag{5.6.24}$$

In this calculation we put a cutoff at $z = \epsilon_0$. To express (5.6.24) in terms of the ultraviolet cutoff ϵ in the BCFT we should instead place the cut-off at $z = \epsilon \kappa e^{\kappa y}$. In the BCFT, this rescaling of the cutoff is related to the Weyl anomaly.

Consider the subregion on the boundary $A_L \cup A_R$, where A_L and A_R are identical copies of the interval $[a, b]$ in the L and R BCFTs. There is some overlap in this section with [121], where they considered the case that A_L and A_R are instead identical copies of the half line $[a, \infty]$. There are two RT surfaces for the region $A_L \cup A_R$ which have both endpoints on the boundary. The first RT surface is a geodesic between $(u_L, y_L, z) = (-u, b, 0)$ and $(u_L, y_L, z) = (-u, a, 0)$ and a geodesic γ_1 between $(u_R, y_R, z) = (u, a, 0)$ and $(u_R, y_R, z) = (u, b, 0)$ which gives

$$\begin{aligned} \frac{A(\gamma_1)}{4G} &= \frac{2c}{3} \log \left(\frac{\beta}{\pi \epsilon} \sinh \frac{\pi}{\beta}(b - a) \right) \\ &\approx \frac{2\pi c}{3\beta}(b - a). \end{aligned} \tag{5.6.25}$$

In going to the second line we used the asymptotic formula for high temperatures and dropped the dependence on cutoff ϵ. The second RT surface is a geodesic between $(u_L, y_L, z) = (-u, b, 0)$ and $(u_R, y_R, z) = (u, b, 0)$ and a geodesic between $(u_L, y_L, z) = (-u, a, 0)$ and $(u_R, y_R, z) = (u, a, 0)$ which gives

$$\begin{aligned} \frac{A(\gamma_2)}{4G} &= \frac{2c}{3} \log \left(\frac{\beta}{\pi \epsilon} \cosh \frac{2\pi}{\beta} u \right) \\ &\approx \frac{4\pi c}{3\beta} u. \end{aligned} \tag{5.6.26}$$

These RT surfaces precisely reproduce the contributions to the entropy (5.3.25) in BCFT which arise from the connected channels (1) and (2) in figure 5.9.

5.6.5 Endpoint on the brane

Now consider the possibility that one of the endpoints of the geodesic lies on the EOW brane. The length of the curve is

$$L = \int_0^1 \frac{\sqrt{-\dot{t}(s)^2 + \dot{x}(s)^2 + \dot{z}(s)^2}}{z(s)} \mathrm{d}s. \tag{5.6.27}$$

where s is a parameter along the curve. The variation is

$$\delta L = \int_0^1 \mathrm{eom}\, \mathrm{d}s + \left. \frac{-\dot{t}(s)\delta t + \dot{x}(s)\delta x + \dot{z}(s)\delta z}{z(s)\sqrt{-\dot{t}(s)^2 + \dot{x}(s)^2 + \dot{z}(s)^2}} \right|_0^1. \tag{5.6.28}$$

The bulk contribution is the equation of motion (eom) for the curve and simply tells us that the curve is a spacelike geodesic, so is given by semicircle. The plane in which it lies depends on where the endpoint of the geodesic intersects the brane. To determine this we need to look at the boundary term. For the endpoint on the EOW brane we impose the Neumann boundary condition[7]

$$-\frac{\mathrm{d}t}{\mathrm{d}s}\delta t + \frac{\mathrm{d}x}{\mathrm{d}s}\delta x + \frac{\mathrm{d}z}{\mathrm{d}s}\delta z = 0. \tag{5.6.29}$$

On the EOW brane the variation of t, x and z are not all independent since

$$-t\delta t + x\delta x + (z + \tan\theta)\delta z = 0. \tag{5.6.30}$$

Hence, the boundary condition (5.6.29) can be written as

$$-\left(x\frac{\mathrm{d}t}{\mathrm{d}s} - t\frac{\mathrm{d}x}{\mathrm{d}s}\right)\delta x + \left(t\frac{\mathrm{d}z}{\mathrm{d}s} - (z + \tan\theta)\frac{\mathrm{d}t}{\mathrm{d}s}\right)\delta z = 0. \tag{5.6.31}$$

We then require the term proportional to δx and δz to vanish independently. The term proportional to δx is a conserved quantity along the geodesic associated to boosts in the (t, x) plane which are generated by the Killing vector field

$$K = \kappa\left(x\frac{\partial}{\partial t} + t\frac{\partial}{\partial x}\right). \tag{5.6.32}$$

This implies

$$x\frac{\mathrm{d}t}{\mathrm{d}s} = t\frac{\mathrm{d}x}{\mathrm{d}s} \tag{5.6.33}$$

holds along the geodesic, not just on the EOW brane. The solution to this equation, $t(s) \propto x(s)$, implies that the endpoint on the brane must lie outside the horizon. This

[7]For the endpoint on the boundary $z = 0$ we impose a Dirichlet boundary condition so that boundary term vanishes.

equation is equivalent to the condition that the endpoint on the brane lies on the same surface of constant u as the endpoint on the boundary.

Requiring the term proportional to δz in (5.6.31) vanishes implies that the geodesic is orthogonal to the EOW brane where they intersect. This is simplest to see in (u, y) coordinates, where the EOW brane is the surface

$$e^{2\kappa y_R} + (z + \tan\theta)^2 = \sec^2\theta. \tag{5.6.34}$$

We can write the two tangent vectors T_1 and T_2 to the this surface as

$$T_1 = \frac{\partial}{\partial u}, \qquad T_2 = \frac{\partial}{\partial y} - \frac{\kappa e^{2\kappa y}}{z + \tan\theta}\frac{\partial}{\partial z}. \tag{5.6.35}$$

In terms of (u, y) coordinates the boundary conditions are

$$\frac{\mathrm{d}u}{\mathrm{d}s} = 0, \qquad \left(\frac{\mathrm{d}z}{\mathrm{d}s} - \kappa(z + \tan\theta)\frac{\mathrm{d}y}{\mathrm{d}s}\right)\Big|_{\mathrm{EOW}} = 0. \tag{5.6.36}$$

We then have

$$(X, T_1) = 0, \qquad (X, T_2) = -\frac{\kappa e^{2\kappa y}}{z + \tan\theta}\left(\frac{\mathrm{d}z}{\mathrm{d}s} - \kappa(z + \tan\theta)\frac{\mathrm{d}y}{\mathrm{d}s}\right), \tag{5.6.37}$$

where X is the tangent to the geodesic. Hence, it follows from the boundary conditions (5.6.36) that X is orthogonal to T_1 and T_2 on the EOW brane.

Let p be the endpoint $(u_R, y_R, z) = (u, y, 0)$ of the geodesic on the boundary and q be the endpoint $(u_R, y_R, z) = (u, y_b, z_b)$ on the brane. The boost isometry maps p and q to the points $(u_R, y_R, z) = (0, y, 0)$ and $(u_R, y_R, z) = (0, y_b, z_b)$ on the $t = 0$ surface, where the geodesic is

$$t = 0, \qquad z = \sqrt{r^2 - (x - x_0 + r)^2}, \tag{5.6.38}$$

where $x_0 = e^{\kappa y}$. Using basic geometry we find[8]

$$r = \sinh\kappa y, \tag{5.6.39}$$

and also

$$z_b = \frac{\cos\theta}{\coth\kappa y + \sin\theta}. \tag{5.6.40}$$

The length of the geodesic is

$$\begin{aligned}\ell(p, q) &= \log\frac{2r}{\epsilon_0} + \tanh^{-1}\sqrt{1 - \frac{z_b^2}{r^2}} + O(\epsilon_0^2) \\ &= \log\frac{2e^{\kappa y}\sinh\kappa y}{\epsilon_0} + \tanh^{-1} T.\end{aligned} \tag{5.6.41}$$

[8] This is explained in figure 5 of [121] where they consider the case that A_L and A_R are identical copies of the half line.

The second term is related to the boundary entropy in the BCFT (see (5.6.10)). There is only one relevant RT surface which involves geodesics ending on the brane. It is a geodesic between $(u_L, y_L, z) = (-u, b, 0)$ and $(u_R, y_R, z) = (u, b, 0)$, a geodesic between $(u_L, y_L, z) = (-u, a, 0)$ and a point on the brane and a geodesic between $(u_R, y_R, z) = (u, a, 0)$ and a point on the brane. The entropy is

$$\begin{aligned}\frac{A(\gamma_3)}{4G} &= 2\log g_b + \frac{c}{3}\log\left(\frac{\beta}{\pi\epsilon}\cosh\frac{2\pi}{\beta}u\right) + \frac{c}{3}\log\left(\frac{\beta}{\pi\epsilon}\sinh\frac{2\pi}{\beta}a\right) \\ &\approx 2\log g_b + \frac{2\pi c}{3\beta}(u+a).\end{aligned} \tag{5.6.42}$$

This RT surface precisely reproduces the contributions to the entropy (5.3.25) in BCFT which arise from the disconnected channel (3) in figure 5.9. Using Takayangi's formula, the entropy of $A_L \cup A_R$ is given by

$$S(A_L \cup A_R) \approx \min\left\{\frac{4\pi c}{3\beta}u, 2\log g_b + \frac{2\pi c}{3\beta}(u+a), \frac{2\pi c}{3\beta}(b-a)\right\}. \tag{5.6.43}$$

Here we wrote the formula for the entropy in the high temperature limit. Notably, according to Takayangi's formula, the dip is absent for large c holographic BCFTs.

5.7 Conclusion and outlook

In BCFT, the boundary acts as a mirror that reflects incoming modes. For spatial intervals of characteristic length L, the thermodynamic limit $L/\beta \gg 1$ uncovers behaviour that is captured by competing OPE channels, and correlations on short scales of order β are invisible. The only effect of the boundary degrees of freedom is to contribute an additive boundary entropy $\log g_b$ in the relevant OPE channels. The free fermion theory is particularly simple given that $\log g_b = 0$. Despite the simplicity of this limit, the time evolution of measures of entanglement can be nontrivial.

These generic features are reproduced by an eternal black hole in AdS2 with matter in the Hartle-Hawking state, in the limit of large intervals such that the scrambling time scale can be neglected, $L \gg \frac{\beta}{2\pi}\log\frac{S_{\rm BH}-S_0}{c}$. If we also take $L \gg k$, we recover the free fermion results with vanishing boundary entropy. In this limit, we expect that the conditions for information recovery to be simple. Since its entropy is much smaller than the thermodynamic entropy of the intervals and the effects of scrambling being negligible, the black hole behaves like an information mirror. The reflected modes on either side of the thermofield double play the role of the Hawking radiation, and the BCFT images on the left and right play the role of the interior of the black hole.

It's remarkable to see the nontrivial dynamical entanglement structure being reproduced by a single island and the qualitative features continuing to apply away from the

strict BCFT limits. A noteworthy outcome of the precise comparison between BCFT and gravity is the identification of different disconnected BCFT channels, corresponding to singularities in the Lorentzian correlator, with bulk channels in JT gravity in the presence of QESs inside or outside the horizon. From the examples studied in this paper, and in the limit of large intervals and small scrambling times, we infer that the QESs serve to reproduce contractions involving BCFT image points in either copy of the TFD state. It would be interesting to perform a more exhaustive study of all possible singularity channels in the BCFT picture (for our two interval problem there are 24 such channels) and the precise characterisation of their associated QESs. It's interesting to note that the purification of modes and consequent dip in entanglement entropy always involves QES behind the horizon. This is similar to the nonequilibrium situation with evaporating black holes [55] where the QES remains inside the horizon while the entropy relaxes after Page time and only pops out of the horizon at parametrically late times when the system is approaching equilibrium.

An interesting aspect of our analysis is that the QESs evolve smoothly across the different regimes which we have identified, including when they plunge into the horizon, the boundaries between the regimes becoming sharp only in the limit of large intervals. In this sense, the JT model appears different to large c holographic BCFTs where the distinct OPE channels are captured by distinct RT surfaces, as we saw in section 5.6. Also important here is that the JT gravity analysis in this paper captures the physics of CFTs with a free quasiparticle description which is distinct from large c holographic BCFTs that are maximally scrambling and are not expected to exhibit a dip in the entanglement entropy. However, the JT gravity plus free fermion setup also allows the exploration of the strong scrambling regime, where it is the black hole that does the scrambling instead of the matter CFT. In particular, it would be interesting to relax the limit of high temperatures and small scrambling times to understand both the deviations from the BCFT picture and the effects of black hole scrambling on the entanglement evolution.

Chapter 6

Moments of the spectral form factor in SYK

This chapter consists of work in progress with S. Prem Kumar and A. Legramandi.

6.1 Introduction

Random matrix universality asserts that correlations between nearby energy levels in chaotic quantum systems are statistically equivalent to those of a random matrix drawn from a Gaussian ensemble [122]. A useful diagnostic of quantum chaos is the spectral form factor (SFF)

$$Z(\mathrm{i}T)Z(-\mathrm{i}T) = \sum_{m,n=1}^{D} e^{-\mathrm{i}T(E_m - E_n)}. \tag{6.1.1}$$

After decaying for a while, the correlations between nearby energy levels shine through at late times: the mean signal follows a linear ramp before saturating on a plateau, with erratic oscillations around the mean of the same order as the signal. Though random matrix universality doesn't describe the noise itself, it does describe the statistics of the noise, as measured by quantities like the variance and higher moments of the SFF.

The aim of this chapter is to study the statistics of the late time noise in the SFF in the SYK model [45, 46]. The SYK model is quantum mechanical model of N Majorana fermions with random all-to-all interactions. The Hamiltonian is given by

$$H = \mathrm{i}^{q/2} \sum_{1\leq i_1<\cdots<i_q\leq N} J_{i_1\ldots i_q}\psi_{i_1}\ldots\psi_{i_q}, \qquad \{\psi_i,\psi_j\} = \delta_{ij}, \tag{6.1.2}$$

where the independent couplings $J_{i_1\ldots i_q}$ are drawn from a Gaussian distribution. After averaging over the couplings, the model can be studied using bilocal collective fields

$G(t,t')$ and $\Sigma(t,t')$. The field $\Sigma(t,t')$ is a Lagrange multiplier enforcing the relation $G(t,t') = \frac{1}{N}\sum_{i=1}^{N}\psi_i(t)\psi_i(t')$. The SYK model came into the spotlight when it was shown that the infrared, the collective field descriptions is governed by the Schwarzian theory, which describes JT gravity [45, 123]. However, in the large N limit, the model is tractable even away from the infrared. The path integral is dominated by the saddle points of the action for the collective fields and small fluctuations around them.

We are now in a position to state the aim of this work more precisely: to understand the statistics of the late time noise in the SFF using the *collective field variables*. These variables are most closely related to the bulk gravitational variables. In this work we will primarily focus on the ramp region. To study the noise, we consider the moments of the spectral form factor $\langle|Z(\mathrm{i}T)|^{2k}\rangle$, where the angled brackets denote an average over the random couplings. To study this quantity in SYK we require a matrix of collective fields $G_{ab}(t,t')$ and $\Sigma_{ab}(t,t')$, where a and b label the $2k$ replicas. Saddle point configurations with nonzero off-diagonal elements represent correlations between the replicas. These saddle points are the collective field analogue of a spacetime wormhole. The case that $k = 1$ was considered in [34].

As a consequence of large D factorisation, random matrix theory (RMT) has a particularly simple prediction for the moments during the ramp region

$$\langle|Z(\mathrm{i}T)|^{2k}\rangle \approx k!\langle|Z(\mathrm{i}T)|^{2}\rangle^{k}. \tag{6.1.3}$$

We trace this behaviour in the collective field variables to a pattern of replica symmetry breaking. More precisely, (6.1.3) is the prediction for systems whose spectra does not have a *mirror symmetry* $E \to -E$. This is the case for the SYK model with $q = 0$ mod 4. For systems with a mirror symmetry, like the SYK model with $q = 2$ mod 4, RMT instead predicts a $(2k-1)!!$ in place of the $k!$. This is reflected in a q mod 4 dependence of the symmetry breaking pattern.

Though the SFF in SYK is not self averaging in the ramp region, in the strict large N limit, its probability distribution is only governed by its mean value: for $q = 0$ mod 4 the distribution is that of the modulus squared of a complex Gaussian variable with zero mean, while for $q = 2$ mod 4 the distribution is that of the square of a real Gaussian variable with zero mean. On the other hand, by including $1/N$ corrections from perturbative fluctuations around the saddle point we find

$$\frac{\langle|Z(\mathrm{i}T)|^{2k}\rangle}{\langle|Z(\mathrm{i}T)|^{2}\rangle^{k}} \approx k!\left[1 + k(k-1)\frac{q!}{N^{q-2}}\left(\frac{E(T)T}{q}\right)^{2} + \dots\right], \tag{6.1.4}$$

for some function $E(T)$ which we do not evaluate exactly. This indicates that for any large but finite N, the distribution of the SFF is no longer so simple. The deviation of the moments becomes significant for sufficiently large k, that is, when k approaches a fixed fraction of $N^{q/2-1}$.

The formula (6.1.4) also hints that the $q = 2$ model is qualitatively different since the $1/N$ correction becomes of order one. This makes sense because the $q = 2$ model

is argued to be integrable unlike its $q > 2$ cousins which are chaotic. By using an analogue of the collective field variables for the $q = 2$ model we show that the pattern of replica symmetry breaking is again the mechanism which explains the noise in the SFF. In contrast to chaotic systems, where the noise in the SFF is typically the same order as the signal, for the $q = 2$ model, the magnitude of the noise is incredibly larger, around $N^{\#T}$ times larger for some number $\#$. The case that $k = 1$ with $q = 2$ was considered in [124].

The plan of this chapter is as follows. We begin with a warm up in section 6.2 by considering a zero-dimensional version of the SYK model called SYK with one time point [125]. Section 6.3 starts with a review of calculation of the SFF in the SYK model [34]. We then proceed to discuss the saddle point configurations for the moments and compute some $1/N$ corrections around the saddle point. In section 6.4 we study the $q = 2$ SYK model which has to be treated separately and highlight the stark difference between the behaviour of the moments in the two models. The first correction to (6.1.4) is inversely proportional to the number of independent random variables in the model. We investigate to what extent this is a generic feature in section 6.5 by considering sparse SYK, where we can tune this variable. Finally, in section 6.6 we end with some conclusions and discuss some directions for further work.

6.2 SYK with one time point

As a warm up to the SYK model, in this section, we consider the finite-dimensional path integral over N Grassmann variables

$$z = \int \mathrm{d}^N \psi \exp\left\{ \mathrm{i}^{q/2} \sum_{i_1 < \cdots < i_q} J_{i_1 \ldots i_q} \psi_{i_1} \ldots \psi_{i_q} \right\}, \tag{6.2.1}$$

where the independent components of the completely antisymmetric tensor of couplings $J_{i_1 \ldots i_q}$ are drawn from a Gaussian distribution with zero mean and variance

$$\langle J_{i_1 \ldots i_q} J_{j_1 \ldots j_q} \rangle = \frac{(q-1)!}{N^{q-1}} \delta_{i_1 j_1} \ldots \delta_{i_q j_q}. \tag{6.2.2}$$

Naturally this model was called SYK with one time point in [125] as it can be thought of as a zero-dimensional version of the SYK model, where the time contour has been shrunk down to a single instant of time. We assume that q is an even integer and N is divisible by q, otherwise z vanishes.

6.2.1 $\langle z^2 \rangle$

Since the average of z vanishes, the simplest nontrivial quantity is $\langle z^2 \rangle$. Doing the Gaussian integral over the couplings and anticommuting the Grassmann variables past each other we find

$$\langle z^2 \rangle = \int \mathrm{d}^{2N}\psi \exp\left\{ \frac{N}{q} \left(\frac{1}{N} \sum_i \psi_i^L \psi_i^R \right)^q \right\}. \tag{6.2.3}$$

This quantity can be represented as an integral over *collective fields*

$$\langle z^2 \rangle = \int_{\mathbb{R}} \mathrm{d}G \int_{\mathrm{i}\mathbb{R}} \frac{\mathrm{d}\Sigma}{2\pi \mathrm{i}/N} \exp\left\{ N \left(\log(\Sigma) - \Sigma G + \frac{1}{q} G^q \right) \right\}. \tag{6.2.4}$$

Here Σ is a Lagrange multiplier enforcing the relation

$$G = \frac{1}{N} \sum_{i=1}^{N} \psi_i^L \psi_i^R. \tag{6.2.5}$$

In contrast to the regular SYK model, where one requires a matrix of collective fields, in this simple model only G_{LR}, which we denote simply by G, is necessary. As we review in the next section, the integral (6.2.4) can be computed exactly [125]. The result is

$$\langle z^2 \rangle = \frac{N!(N/q)^{N/q}}{N^N (N/q)!}. \tag{6.2.6}$$

For large N the integral can also be computed using the saddle point approximation. The saddle point equation is

$$G = \frac{1}{\Sigma}, \qquad \Sigma = G^{q-1}. \tag{6.2.7}$$

There are q solutions, each corresponding to a q^{th} root of unity and contributing as

$$\frac{1}{\sqrt{q}} e^{-\left(1-\frac{1}{q}\right)N} e^{2\pi \mathrm{i} m N/q}, \qquad m = 0, \ldots, q-1. \tag{6.2.8}$$

Summing over the q saddle points gives

$$\langle z^2 \rangle \sim \sqrt{q} e^{-\left(1-\frac{1}{q}\right)N}, \tag{6.2.9}$$

which reproduces the large N asymptotic form of the exact result (6.2.6). We will occasionally refer to these saddle points as *wormholes* as they represent a nonzero correlation between the L and R replicas.

6.2.2 $\langle z^{2k} \rangle$

Repeating the G, Σ trick for $\langle z^{2k} \rangle$ we find[1]

$$\langle z^{2k} \rangle = \int_{\mathbb{R}} \mathrm{d}G_{ab} \int_{\mathrm{i}\mathbb{R}} \frac{\mathrm{d}\Sigma_{ab}}{2\pi\mathrm{i}/N} \mathrm{Pf}(\Sigma)^N \exp\left\{ N \sum_{a<b} \left(-\Sigma_{ab} G_{ab} + \frac{1}{q} G_{ab}^q \right) \right\}, \tag{6.2.10}$$

where Pf is the Pfaffian, and G_{ab} and Σ_{ab} are $2k \times 2k$ antisymmetric matrices. Using the integral representation of the delta function and the definition of the Pfaffian, the integral over Σ_{ab} can be evaluated exactly

$$\int_{\mathrm{i}\mathbb{R}} \frac{\mathrm{d}\Sigma_{ab}}{2\pi\mathrm{i}/N} \mathrm{Pf}(\Sigma)^N \exp\left\{ -N \sum_{a<b} \Sigma_{ab} G_{ab} \right\} = N^{-kN} \mathrm{Pf}(\partial_G)^N \delta(G). \tag{6.2.11}$$

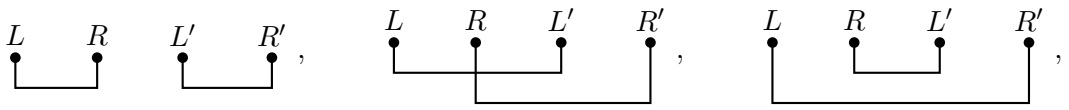
This leaves us with

$$\langle z^{2k} \rangle = N^{-kN} \mathrm{Pf}(\partial_G)^N \exp\left\{ N \sum_{a<b} G_{ab}^q \right\} \Bigg|_{G_{ab}=0}. \tag{6.2.12}$$

Using the fact that the Pfaffian can be written as a sum over pairings we find

$$\langle z^{2k} \rangle = N^{-kN} \sum_{n_\pi} \frac{N!}{\prod_{\pi \in P_{2k}} n_\pi!} \prod_{a<b} \left(\frac{N}{q} \right)^{m_{ab}} \frac{(q m_{ab})!}{m_{ab}!}, \qquad q m_{ab} = \sum_{\substack{\pi \in P_{2k}: \\ (a,b) \in \pi}} n_\pi, \tag{6.2.13}$$

where P_{2k} denotes the set of pairings of $2k$ elements. In this expression the first sum is over integers $n_\pi \geq 0$ such that both $\sum_{\pi \in P_{2k}} n_\pi = N$ and m_{ab} is an integer. For $k = 2$ this formula simplifies: there are 3 parameters n_1, n_2, and n_3 which correspond to the 3 pairings,



and 6 parameters m_{ab},

$$q m_{LR} = q m_{L'R'} = n_1, \quad q m_{LL'} = q m_{RR'} = n_2, \qquad q m_{LR'} = q m_{RL'} = n_3. \tag{6.2.14}$$

Since the n_i must be divisible by q we can rescale $n_i \to q n_i$ so that the sum is instead over integers $n_i \geq 0$ such that $n_1 + n_2 + n_3 = N/q$. This gives

$$\langle z^4 \rangle = \frac{N!}{N^2} \left(\frac{N}{q} \right)^{\frac{2N}{q}} \sum_{\substack{n_1+n_2+n_3=N/q: \\ n_i \geq 0}} \frac{(q n_1)!(q n_2)!(q n_3)!}{(n_1)!^2 (n_2)!^2 (n_3)!^2}. \tag{6.2.15}$$

[1] The odd moments of z vanish as the Pfaffian of odd dimensional matrix vanishes.

In the large N limit and when $q > 2$

$$\frac{\langle z^4 \rangle}{\langle z^2 \rangle^2} = 3 + 6 \times \frac{q!}{q^2} \frac{1}{N^{q-2}} + \dots \tag{6.2.16}$$

The leading term arises from terms in the sum where one of the n_i are equal to N/q and the other n_i vanish. The first correction arises from terms in the sum where one of the n_i are equal to $N/q - 1$, another n_i is equal to 1, and the third n_i vanishes. The factors 3 and 6 count the multiplicity of such terms. For general k, in the large N limit and when $q > 2$

$$\frac{\langle z^{2k} \rangle}{\langle z^2 \rangle^k} = (2k-1)!! \left[1 + k(k-1) \times \frac{q!}{q^2} \frac{1}{N^{q-2}} + \dots \right]. \tag{6.2.17}$$

The leading term is determined by maximising the number of pairs (a, b) for which m_{ab} takes its maximal value, which is N/q. The maximum number of such pairs is k. This occurs when one of the n_i are equal to N and the other n_i vanish. Since there are $|P_{2k}| = (2k-1)!!$ such terms this leads to the contribution

$$\langle z^{2k} \rangle \supset (2k-1)!! \left[\frac{N!(N/q)^N}{N^N (N/q)!} \right]^k. \tag{6.2.18}$$

The first correction arises from a subset of the terms in the sum where one of the n_i are equal to $N - q$, another n_i is equal to q, and the remaining n_i vanish. Let ℓ denote the number of pairs in common between the pairing with n_i equal to $N - q$ and the pairing with n_i equal to q. Then ℓ of the parameters m_{ab} are equal to N/q, $k - \ell$ of parameters m_{ab} are equal to $N/q - 1$, another $k - \ell$ of parameters m_{ab} are equal to 1, and the remaining parameters vanish. The dominant contribution comes from the terms with $\ell = k - 2$,

$$\langle z^{2k} \rangle \supset (2k-1)!! \times k(k-1) \frac{(N/q)^{kN/q} N!^{k-1} (N-q)! q!}{N^{kN} (N/q)!^{k-2} (N/q - 1)!^2}. \tag{6.2.19}$$

The combinatorial factor $(2k-1)!! \times k(k-1)$ can be understood as follows. For a fixed pairing we need to sum over the number of ways to pick another pairing such that the two pairings have $k-2$ pairs in common. There are $k(k-1)$ ways to do this. Summing over the number of ways of picking the first pairing that we fixed gives the factor $|P_{2k}| = (2k-1)!!$.

We learn from (6.2.17) that, in the strict large N limit, z approaches a Gaussian variable with zero mean and variance given by (6.2.9), since $(2k-1)!!$ are the moments of the Gaussian distribution. On the other hand, for any large but finite N, the structure of the $1/N$ corrections in (6.2.17) indicate that the distribution of z is no longer Gaussian. The deviation of the moments becomes significant for sufficiently large k, that is, when k approaches a fixed fraction of $N^{q/2-1}$. The formula (6.2.17) also hints that the $q = 2$ model is qualitatively different since the $1/N$ correction becomes of order one. This is indeed the case and the $q = 2$ model needs to be treated separately. We will not discuss this model here as it's discussed in detail in [125].

6.2.3 $\langle z^{2k}\rangle$ from saddle points

As before, the leading behaviour (6.2.18) of $\langle z^{2k}\rangle$ for large N can also be obtained from the saddle point approximation. There are $(2k-1)!! \times q^k$ particularly simple saddle points which represent pairings between the $2k$ replicas. Concretely, to each pairing $\pi \in P_{2k}$ of the $2k$ replicas set $G_{ab} = \Sigma_{ab} = 0$ if $(a,b) \notin \pi$. The saddle point equations then reduce to k decoupled equations

$$G_{ab} = \frac{1}{\Sigma_{ab}}, \qquad \Sigma_{ab} = G_{ab}^{q-1}, \tag{6.2.20}$$

for the remaining variables G_{ab} and Σ_{ab} with $(a,b) \in \pi$. This is nothing but k copies of the saddle point equation (6.2.7) for $\langle z^2\rangle$. For each pair $(a,b) \in \pi$ there are q solutions. Therefore, for each pairing $\pi \in P_{2k}$ there are q^k solutions. Since there are $(2k-1)!!$ possible pairings, the total number of saddle points is $(2k-1)!! \times q^k$. Summing over the saddle points reproduces the large N asymptotic form (6.2.18) of the exact result. Actually, to claim that we get (6.2.18) we need to do a one loop computation. We will explain how this works shortly.

Another way to understand the factor $(2k-1)!!$ is the following. The collective field representation (6.2.10) of $\langle z^{2k}\rangle$ has a discrete symmetry S_{2k}/S_2^k which acts by permuting the $2k$ replicas. The quotient by S_2^k ensures that we get an action on the collective fields, which are constrained to be antisymmetric. The factor $(2k-1)!!$ then corresponds to the size of the orbit of a single *wormhole pairing* under the action of S_{2k}/S_2^k. This space is the quotient $S_{2k}/(S_k \times S_2^k)$, where S_k is the subgroup of S_{2k} which leaves the wormhole pairing fixed.

It turns out the first correction (6.2.19) to $\langle z^{2k}\rangle/\langle z^2\rangle^k$ can be computed by considering perturbative fluctuations around the wormhole pairing saddle points. To see this, we expand in small fluctuations around one of the $(2k-1)!!$ saddle points configurations G_{ab} and $\mathsf{\Sigma}_{ab}$,

$$G_{ab} = \mathsf{G}_{ab} + \delta G_{ab}, \qquad \Sigma_{ab} = \mathsf{\Sigma}_{ab} + \delta\Sigma_{ab}. \tag{6.2.21}$$

The action decomposes as $I = I_{\rm cl} + \delta I$, where $I_{\rm cl}$ is the classical action, and

$$\frac{1}{N}\delta I = \sum_{n\geq 2}\frac{(-1)^n}{2n}\mathrm{Tr}[(\mathsf{G}\delta\Sigma)^n] + \sum_{a<b}\left[\delta\Sigma_{ab}\delta G_{ab} - \frac{1}{q}\sum_{n=2}^{q}\binom{q}{2}\mathsf{G}_{ab}^{q-n}\delta G_{ab}^n\right]. \tag{6.2.22}$$

In obtaining this expression we used the saddle point equations (6.2.20). Let's begin by computing the one loop contribution, for which we only keep the quadratic part of the action

$$\frac{1}{N}\delta I \supset \frac{1}{4}\mathrm{Tr}[(\mathsf{G}\delta\Sigma)^2] + \sum_{a<b}\left[\delta\Sigma_{ab}\delta G_{ab} - \frac{q-1}{2}\mathsf{G}_{ab}^{q-2}\delta G_{ab}^2\right]. \tag{6.2.23}$$

Any $2k \times 2k$ antisymmetric matrix M can be decomposed as $M = M_{\parallel} + M_{\perp}$, where $M_{\parallel}$ lies in the subspace spanned by $2k \times 2k$ antisymmetric matrices of the form

$$\begin{array}{c} \begin{array}{cccccccc} L_1 & R_1 & L_2 & R_2 & \dots & L_k & R_k & \end{array} \\ \begin{pmatrix} 0 & * & & & & & & \\ * & 0 & & & & & & \\ & & 0 & * & & & & \\ & & * & 0 & & & & \\ & & & & \ddots & & & \\ & & & & & 0 & * & \\ & & & & & * & 0 & \end{pmatrix} \end{array} \begin{array}{c} L_1 \\ R_1 \\ L_2 \\ R_2 \\ \vdots \\ L_k \\ R_k \end{array} \tag{6.2.24}$$

and $M_{\perp}$ lies in its orthocomplement. If G and $\mathsf{\Sigma}$ represent the saddle point which takes the block diagonal form (6.2.24) then $\mathsf{G} = \mathsf{G}_{\parallel}$ and $\mathsf{\Sigma} = \mathsf{\Sigma}_{\parallel}$. Equipped with this notation, the quadratic part of the action may be written as

$$\frac{N}{4}\mathrm{Tr}[(\mathsf{G}\delta\Sigma)^2] + N\sum_{a<b}\left[(\delta\Sigma_{\parallel})_{ab}(\delta G_{\parallel})_{ab} + (\delta\Sigma_{\perp})_{ab}(\delta G_{\perp})_{ab} - \frac{q-1}{2}\mathsf{G}_{ab}^{q-2}(\delta G_{\parallel})_{ab}^2\right]. \tag{6.2.25}$$

Notice that $\delta G_{\perp}$ appears linearly and so integrating it out gives a delta function which sets $\delta\Sigma_{\perp} = 0$. This leaves us with an integral over the remaining variables $\delta G_{\parallel}$ and $\delta\Sigma_{\parallel}$ with quadratic action

$$\frac{N}{4}\mathrm{Tr}[(\mathsf{G}\delta\Sigma_{\parallel})^2] + N\sum_{a<b}\left[(\delta\Sigma_{\parallel})_{ab}(\delta G_{\parallel})_{ab} - \frac{q-1}{2}\mathsf{G}_{ab}^{q-2}(\delta G_{\parallel})_{ab}^2\right], \tag{6.2.26}$$

which factorises into a sum over the k blocks in (6.2.24). This implies that the one loop contribution from expanding in fluctuations around one of the $(2k-1)!!$ saddle points for $\langle z^{2k}\rangle$ is given by the k^{th} power of the one loop contribution for $\langle z^2\rangle$. This proves that the leading behaviour for large N

$$\frac{\langle z^{2k}\rangle}{\langle z^2\rangle^k} = (2k-1)!!. \tag{6.2.27}$$

Although an excellent approximation for large N, we will see that this simple relation eventually breaks down for large enough k after including $1/N$ corrections. These corrections can be computed using Feynman diagrams together with the free propagators

$$\begin{aligned} \langle \delta G_{ab}\delta G_{cd}\rangle &= \mathsf{G}_{ac}\mathsf{G}_{bd} - \mathsf{G}_{ad}\mathsf{G}_{bc} - (1-1/q)\mathsf{G}_{ab}^2\delta_{ac}\delta_{bd}, \\ \langle \delta G_{ab}\delta\Sigma_{cd}\rangle &= 1 - (1-1/q)\pi_{ab,cd}, \\ \langle \delta\Sigma_{ab}\delta\Sigma_{cd}\rangle &= -(1-1/q)\mathsf{G}_{ab}^{q-2}\delta_{ac}\delta_{bd}, \end{aligned} \tag{6.2.28}$$

obtained from the quadratic part of the action (6.2.23). In this expression the angular brackets represent an average weighted by the quadratic action (6.2.25) and π is the projector onto the subspace of $2k \times 2k$ antisymmetric matrices of the form (6.2.24). Notice that the only interaction term in δI which involves $\delta G_\perp$ is proportional to δG^q_{ab}. We can then split the Feynman diagrams we encounter into two types: diagrams with no $\delta G^q_\perp$ vertices and diagrams with at least one $\delta G^q_\perp$ vertex. For the first type of diagrams we can repeat the argument given above and integrate out $\delta G_\perp$ giving a delta function which sets $\delta\Sigma_\perp = 0$. Summing up these terms precisely gives the perturbative contribution of $\langle z^2\rangle^k$. It's clear then that corrections to the leading order behaviour of the ratio $\langle z^{2k}\rangle/\langle z^2\rangle^k$ come from Feynman diagrams with at least one $\delta G^q_\perp$ interaction vertex.

The simplest Feynman diagram to write down involves just one $\delta G^q_\perp$ vertex. However, using (6.2.28), we find these contributions vanish

$$\frac{N}{q}\sum_{a<b}\langle(\delta G_\perp)^q_{ab}\rangle = 0. \tag{6.2.29}$$

Another simple diagram to write down involves one $\delta G^q_\perp$ and one δG^q vertex. One might think that there could be a diagram with one $\delta G^q_\perp$ vertex and one δG^n vertex with $n < q$, however, by similar a similar calculation to the one above, this term necessarily vanishes too. Summing up diagrams with one $\delta G^q_\perp$ vertex and one δG^q vertex gives

$$\frac{N^2}{2q^2}\sum_{\substack{a<b\\c<d}}\langle(\delta G_\perp)^q_{ab}\delta G^q_{cd}\rangle. \tag{6.2.30}$$

This can be computed using Wick's theorem and the free propagator (6.2.28). The only nonzero Wick contractions are where a $(\delta G_\perp)_{ab}$ is contracted with a δG_{cd}. Since there are $q!$ such contractions,

$$q!\frac{N^2}{2q^2}\sum_{\substack{a<b\\c<d}}\langle(\delta G_\perp)_{ab}\delta G_{cd}\rangle^q = k(k-1)\frac{q!}{q^2}\frac{1}{N^{q-2}}. \tag{6.2.31}$$

Since this contribution comes from Feynman diagrams with two vertices and q propagators we get an overall contribution of $1/N^{q-2}$. Notice that this precisely reproduces the first correction obtained from the exact formula (6.2.17). A proof of that any other diagrams either vanish or are subleading will be given in the upcoming work.

6.3 The SYK model

In this section we consider the moments of the spectral form factor in the SYK model. The SYK model is a quantum mechanical model of N Majorana fermions $\psi_1, \ldots, \psi_N$

with random couplings. The Hamiltonian is given by

$$H = \mathrm{i}^{q/2} \sum_{1\le i_1<\cdots<i_q\le N} J_{i_1\ldots i_q}\psi_{i_1}\ldots\psi_{i_q}, \qquad \{\psi_i,\psi_j\} = \delta_{ij}. \tag{6.3.1}$$

The independent components of the antisymmetric tensor of couplings $J_{i_1\ldots i_q}$ are drawn from a Gaussian distribution with zero mean and variance

$$\langle J_{i_1\ldots i_q} J_{j_1\ldots j_q}\rangle = \frac{J^2(q-1)!}{N^{q-1}}\delta_{i_1 j_1}\ldots\delta_{i_q j_q}. \tag{6.3.2}$$

Throughout we assume both N and q are even. Since N is even, the fermion algebra (6.3.1) has an irreducible representation in a Hilbert space of dimension $D = 2^{N/2}$.

6.3.1 Collective fields

The spectral form factor may be represented as a path integral

$$|Z(\mathrm{i}T)|^2 = \int D\psi \exp\left\{\mathrm{i}\int_0^T \mathrm{d}t\left[\frac{\mathrm{i}}{2}\psi_i^a\partial_t\psi_i^a - J_{i_1\ldots i_q}\left(\mathrm{i}^{\frac{q}{2}}\psi_{i_1}^L\ldots\psi_{i_q}^L - (-\mathrm{i})^{\frac{q}{2}}\psi_{i_1}^R\ldots\psi_{i_q}^R\right)\right]\right\}, \tag{6.3.3}$$

with antiperiodic boundary conditions for the fermions, $\psi_i^a(T) = -\psi_i^a(0)$, around the real time circle parameterised by $t\in[0,T]$. In this expression $a\in\{L,R\}$ is a replica index and the integral over ψ^L computes $\mathrm{Tr}[e^{-\mathrm{i}TH}]$ whilst the integral over ψ^R computes $\mathrm{Tr}[e^{\mathrm{i}TH}]$. The kinetic term involves an implicit sum over both i and a whilst the interaction term involves an implicit sum over $i_1<\cdots<i_q$.

The k^{th} moment of the spectral form factor may be represented as a path integral over *collective fields*

$$\begin{aligned}
&\langle|Z(\mathrm{i}T)|^{2k}\rangle = \int DGD\Sigma\, e^{-I[G,\Sigma]}\\
&\frac{1}{N}I[G,\Sigma] = -\frac{1}{2}\log\det(\partial_t-\Sigma) + \frac{1}{2}\int_0^T\int_0^T \mathrm{d}t\mathrm{d}t'\left[\Sigma_{ab}(t,t')G_{ab}(t,t') - \frac{J^2}{q}s_{ab}G_{ab}(t,t')^q\right]\\
&s_{L_\# L_\#} = s_{R_\# R_\#} = -1, \qquad s_{L_\# R_\#} = s_{R_\# L_\#} = \mathrm{i}^q.
\end{aligned} \tag{6.3.4}$$

Here $\Sigma_{ab}(t,t')$ is a Lagrange multiplier enforcing the relation

$$G_{ab}(t,t') = \frac{1}{N}\sum_{i=1}^N \psi_i^a(t)\psi_i^b(t'). \tag{6.3.5}$$

The indices $a,b\in\{L_1,R_1,\ldots L_k,R_k\}$ label the L and R replicas, which were each introduced to represent the factors of $\mathrm{Tr}[e^{-\mathrm{i}TH}]$ and $\mathrm{Tr}[e^{\mathrm{i}TH}]$ in the spectral form factor. The collective fields G,Σ are antiperiodic with period T in each argument, a property inherited from the antiperiodic boundary conditions on the fermions. Also, not all

components of G and Σ are independent since $G_{ab}(t,t') = -G_{ba}(t',t)$ and $\Sigma_{ab}(t,t') = -\Sigma_{ba}(t',t)$, which follows from the Grassmann nature of the fermions. Finally, in the action (6.3.4), the determinant represents the ordinary determinant in replica space as well as the functional determinant.

The action (6.3.4) has a $U(1)^{2k}$ time translation symmetry which acts independently on each of the L and R systems. In addition, there is a discrete replica symmetry

$$G_{ab} \to G_{\pi(a)\pi(b)}, \qquad \Sigma_{ab} \to \Sigma_{\pi(a)\pi(b)}. \tag{6.3.6}$$

For $q = 0 \bmod 4$, $\pi \in S_k^L \times S_k^R$ permutes the L and R replicas separately, whereas, for $q = 2 \bmod 4$, $\pi \in S_{2k}/S_2^k$ can also permute the L and R replicas amongst each other. The dependence of the replica symmetry on $q \bmod 4$ stems from the fact that exchanging an L and R replica is only a symmetry if $Z(\mathrm{i}T)$ and $Z(-\mathrm{i}T)$ are the same variables. In other words, it is a symmetry if the spectrum has a mirror symmetry $E \leftrightarrow -E$, which is the case for model with $q = 2 \bmod 4$. The signature of this symmetry in the collective field description (6.3.4) lies in the dependence of s_{ab} on $q \bmod 4$.

Assuming the relevant saddle point configurations preserve $U(1)^k_{\text{diag}} \subset U(1)^{2k}$ symmetry, that is, the saddle point configurations only depend on the difference of times, the saddle point equations can written as

$$\begin{aligned} G(\omega_n) &= -\left(\mathrm{i}\omega_n + \Sigma(\omega_n)\right)^{-1}, \\ \Sigma_{ab}(t) &= s_{ab} J^2 G_{ab}(t)^{q-1}. \end{aligned} \tag{6.3.7}$$

In the first equation $G(\omega_n)$ and $\Sigma(\omega_n)$ are the Fourier modes of the collective fields and frequencies are fermionic Matsubara frequencies due to the antiperiodic boundary conditions.

6.3.2 The spectral form factor

In this section we review the calculation of the spectral form factor, that is, the case that $k = 1$, which was analysed by Saad, Shenker, and Stanford in [34]. If we assume the off-diagonal components of the collective fields vanish, that is, $G_{LR} = \Sigma_{LR} = 0$, then the saddle points equations (6.3.7) decouple into separate equations for the L and R systems. These saddle points give rise to the disconnected contribution of the spectral form factor

$$\langle |Z(\mathrm{i}T)|^2 \rangle \supset |\langle Z(\mathrm{i}T) \rangle|^2. \tag{6.3.8}$$

This describes the slope of the spectral form factor and it decays to zero for late times. SSS showed that there also exist saddle points where the off-diagonal components of the collective fields do not vanish. A consequence is that these configurations spontaneously break the relative time translation symmetry. This means that there is a continuous family of saddle points which can be generated by acting on this solution with the

generator of the broken symmetry. Concretely, this symmetry acts on the saddle points by leaving G_{LL} and G_{LR} unchanged but shifting $G_{LR}(t) \to G_{LR}(t-\Delta)$. There is a similar action on Σ. Since the collective fields are antiperiodic with period T, the relative time shift parameter Δ is valued on a circle of circumference $2T$. The volume of the orbit generated by the broken symmetry is simply $2T$ which gives rise to the linear ramp in the spectral form factor.

As argued by SSS, there is also another zero mode for these saddle point configurations which can be understood by considering an auxiliary problem. The idea is to think of computing the thermal partition function

$$Z(\beta_{\text{aux}}) = \text{Tr}[e^{-\frac{\beta_{\text{aux}}}{2}H} e^{-\text{i}TH} e^{-\frac{\beta_{\text{aux}}}{2}H} e^{\text{i}TH}], \tag{6.3.9}$$

by a path integral on an elaborate Schwinger-Keldysh contour. In this problem the field configurations on the two Lorentzian parts of the contour are joined to each other along the Euclidean parts of the contour. In contrast, the original problem of computing $\text{Tr}[e^{-\text{i}TH}]\text{Tr}[e^{\text{i}TH}]$, requires separately periodically identifying the two Lorentzian contours. However, for large T, the saddle point configurations for the Lorentzian parts of the contour of the auxiliary problem are approximately same as that for the original problem. This suggests we can construct an approximate solution for the original problem by taking fermion correlators in the thermofield double state[2],

$$G_{ab}^{(\beta_{\text{aux}})}(t) = \frac{1}{N}\sum_{i=1}^{N} \langle \Psi_{\text{TFD}}| \psi_i^a(t)\psi_i^b(0) |\Psi_{\text{TFD}}\rangle \,, \tag{6.3.10}$$

and summing over images

$$G_{ab}(t) = G_{ab}^{(\beta_{\text{aux}})}(t) - G_{ab}^{(\beta_{\text{aux}})}(t-T). \tag{6.3.11}$$

More precisely, SSS showed numerically that for large T there exists an exact solution to the saddle point equation which lies nearby this approximate solution. These configurations cannot be written down analytically. However, in the infrared limit where the SYK model becomes nearly conformal, we can be more explicit. In this case [34]

$$\begin{aligned} G_{LL}^{(\beta_{\text{aux}})}(t) &= G_{RR}^{(\beta_{\text{aux}})}(t) = b\left[\frac{\beta_{\text{aux}}}{\pi}\sinh\left(\frac{\pi t}{\beta_{\text{aux}}}\right)\right]^{-\frac{2}{q}} \text{sign}(t), \\ G_{LR}^{(\beta_{\text{aux}})}(t) &= \text{i}b\left[\frac{\beta_{\text{aux}}}{\pi}\cosh\left(\frac{\pi(t-\Delta)}{\beta_{\text{aux}}}\right)\right]^{-\frac{2}{q}}, \end{aligned} \tag{6.3.12}$$

where b is a constant determined by $J^2 b^q \pi \cot \pi/q = 1/2 - 1/q$. The upshot of studying this auxiliary problem is that it makes it clear that there is another zero mode, namely

[2] As usual, left-right correlators in the thermofield double state can be obtained from right-right thermal correlators by analytic continuation in t.

β_{aux}. Also, since the classical action for the two Lorentzian parts of the contour in the auxiliary problem ought to cancel out it suggests the action for these saddle points should be zero. SSS gave a more careful argument and showed the action is zero, up to exponentially small corrections in T. Hence, these saddle points contribute as

$$\langle |Z(\mathrm{i}T)|^2 \rangle \supset \int_0^\infty \mathrm{d}\beta_{\text{aux}} \mu(\beta_{\text{aux}}) \int_0^{2T} \mathrm{d}\Delta, \tag{6.3.13}$$

where the function $\mu(\beta_{\text{aux}})$ is determined by a one loop computation. For large T, it turns out that the measure $\mathrm{d}\beta_{\text{aux}}\mu(\beta_{\text{aux}})$ becomes a flat measure in terms of the energy. This gives

$$\langle |Z(\mathrm{i}T)|^2 \rangle \supset 2T \int \frac{\mathrm{d}E_{\text{aux}}}{2\pi} \times \begin{cases} 2 & \text{for } q = 0 \bmod 4, \\ 1 & \text{for } q = 2 \bmod 4. \end{cases} \tag{6.3.14}$$

The model with $q = 0 \bmod 4$ has a time reversal symmetry which leads to an extra factor of two since we need to sum over the saddle point and its time reversal conjugate.

6.3.3 General k

There is a simple class of saddle points which contribute to the k^{th} moment of the spectral form factor. To understand them, consider a configuration where the collective fields G and Σ are block diagonal:

$$\begin{array}{c} \begin{array}{cccccccc} L_1 & R_1 & L_2 & R_2 & \dots & L_k & R_k & \end{array} \\ \begin{pmatrix} * & * & & & & & & \\ * & * & & & & & & \\ & & * & * & & & & \\ & & * & * & & & & \\ & & & & \ddots & & & \\ & & & & & * & * & \\ & & & & & * & * & \end{pmatrix} \end{array} \begin{array}{c} L_1 \\ R_1 \\ L_2 \\ R_2 \\ \vdots \\ L_k \\ R_k \end{array} \tag{6.3.15}$$

On these configurations, the saddle point equations (6.3.7) reduce to k decoupled equations for the blocks. This is nothing but k copies of the saddle point equations for the spectral form factor. For each block there are two simple solutions: a disconnected solution and a connected solution. The disconnected solution refers to the saddle point where the replicas are uncorrelated, that is, $G_{LR} = \Sigma_{LR} = 0$, whilst the connected solution refers to the saddle points found by SSS where where the replicas are correlated, that is, $G_{LR} = \Sigma_{LR} \neq 0$. For the block diagonal configuration there are 2^k saddle points which correspond to taking $k - \ell$ of the blocks of the disconnected solution and

k blocks of the connected solution with ℓ running from 0 to k. Taking all the blocks to be the disconnected solution, that is, $\ell = 0$, gives rise to the disconnected contribution

$$\langle |Z(\mathrm{i}T)|^{2k} \rangle \supset |\langle Z(\mathrm{i}T) \rangle|^{2k}. \tag{6.3.16}$$

In general, the disconnected solutions give rise to a contribution which decays as a power law $T^{-\#(k-\ell)}$ for late times. However, since the solution spontaneously breaks ℓ of the relative time translation symmetries, for nonzero ℓ, this decay is balanced by the volume of the orbit generated by the broken symmetry which is proportional to T^{ℓ}. For late times, the dominant contribution arises from taking all the blocks to be the disconnected solution, that is, $\ell = k$. This gives rise to the contribution

$$\langle |Z(\mathrm{i}T)|^{2k} \rangle \supset \langle |Z(\mathrm{i}T)|^{2} \rangle^{k} \times \begin{cases} k! & \text{for } q = 0 \bmod 4, \\ (2k-1)!! & \text{for } q = 2 \bmod 4. \end{cases} \tag{6.3.17}$$

The combinatorial factors represent the noise around the mean value of the spectral form factor and follow from the pattern of replica symmetry breaking. More specifically, the collective field representation of the k^{th} moment of the spectral form factor has a discrete replica symmetry (6.3.6). The combinatorial factors then arise as the size of the orbit of a single block diagonal solution under the action of the discrete symmetry group. The quotient space is $(S_k \times S_k)/S_k$ for $q = 0$ mod 4 and $S_{2k}/(S_k \times S_2^k)$ for $q = 2$ mod 4. In both cases, S_k is the subgroup which leaves the block diagonal solution fixed. Though the spectral form factor is not self averaging in the ramp region, its probability distribution is only governed by its mean value: for $q = 0$ mod 4, it follows an exponential distribution, while for $q = 2$ mod 4, it follows a chi-squared distribution.

In principle, there could be contributions from other saddle points. In figure 6.1, we compute the ratio $\langle |Z(\mathrm{i}T)|^{2k} \rangle / \langle |Z(\mathrm{i}T)|^{2} \rangle^{k}$ numerically for the SYK model with $q = 4$ and $q = 6$. These results seem to support the idea that other saddle points do not contribute in a relevant way.

6.3.4 1/N corrections

In this section we try to understand the leading correction to the formula (6.3.17) from perturbative fluctuations around the saddle point. The calculation is similar in spirit to the one in section 6.2.3 for SYK with one time point. We expand in small fluctuations around one of the block diagonal saddle point configurations G_{ab} and $\mathsf{\Sigma}_{ab}$, where all the blocks are taken to be the connected solution,

$$G_{ab} = \mathsf{G}_{ab} + \delta G_{ab}, \qquad \Sigma_{ab} = \mathsf{\Sigma}_{ab} + \delta \Sigma_{ab}. \tag{6.3.18}$$

The action decomposes as $I = I_{\text{cl}} + \delta I$, where I_{cl} is the classical action, which is exponentially small in T for large T, and

$$\frac{1}{N}\,\delta I = \sum_{n \geq 2} \frac{1}{2n} \mathrm{Tr}[(\mathsf{G}\delta\Sigma)^n] + \frac{1}{2}\int_0^T \int_0^T \mathrm{d}t\mathrm{d}t' \left[\delta\Sigma_{ab}\delta G_{ab} - \frac{J^2}{q} s_{ab} \sum_{n=2}^{q} \binom{q}{n} \mathsf{G}_{ab}^{q-n} \delta G_{ab}^{n} \right]. \tag{6.3.19}$$

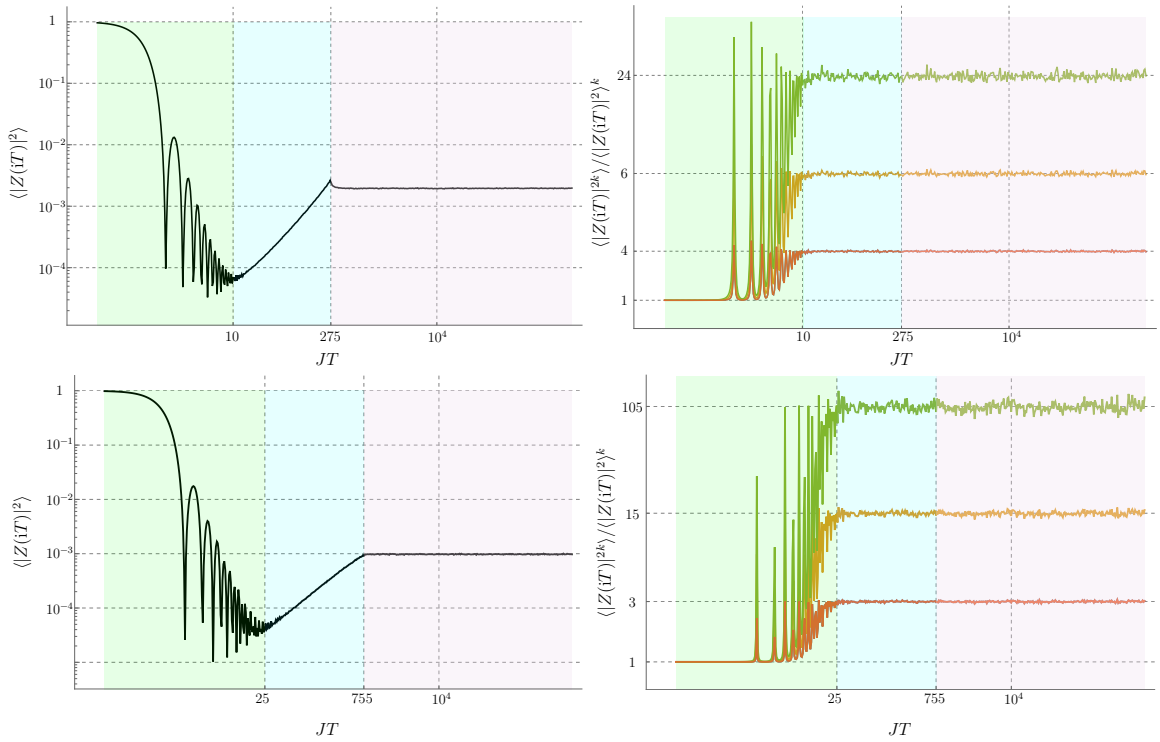


Figure 6.1: The top and bottom panels are for the SYK model with $q = 4$ and $q = 6$, respectively, $N = 20$, and averaged over 2×10^4 realisations. The slope, ramp, and plateau regions are represented as the shaded green, blue, and pink regions, respectively. On the right $\langle |Z(\mathrm{i}T)|^{2k} \rangle / \langle |Z(\mathrm{i}T)|^2 \rangle^k$ is plotted for $k = 2$ (red), $k = 3$ (orange), and $k = 4$ (green). In this figure, g denotes the spectral form factor and we use the abbreviation $\mathsf{SFF}_{2k} = \langle |Z(\mathrm{i}T)|^{2k} \rangle$.

Here we dropped the time arguments but in all cases the fluctuations always depend on a pair of times. Let's begin by computing the one loop contribution, for which we only keep the quadratic part of the action

$$\frac{1}{N}\,\delta I \supset \frac{1}{4}\mathrm{Tr}[(\mathsf{G}\delta\Sigma)^2] + \frac{1}{2}\int_0^T\int_0^T \mathrm{d}t\mathrm{d}t' \left[\delta\Sigma_{ab}\delta G_{ab} - J^2\frac{q-1}{2}s_{ab}\mathsf{G}_{ab}^{q-2}\delta G_{ab}^2\right]. \quad (6.3.20)$$

As before, the fluctuation δG can be decomposed as $\delta G = \delta G_\parallel + \delta G_\perp$, where $\delta G_\parallel$ lies in the subspace spanned by matrices of the form (6.3.15) and $\delta G_\perp$ lies in its orthocomplement. There is a similar decomposition for $\delta\Sigma$. Since $\delta G_\perp$ appears linearly in the action, integrating it out gives a delta function which sets $\delta\Sigma_\perp = 0$. It's then straightforward to see that the one loop contribution factorises over the k blocks. This means that the formula (6.3.17) is indeed correct to leading order in the limit of large N. As explained in section 6.2.3, the $1/N$ corrections to this formula come from Feynman diagrams with at least one $\delta G_\perp^q$ interaction vertex. In this case, things are trickier since we can't find the free propagator exactly as we could in SYK with one time point. However, it turns out that one can understand quite a bit just by studying the structure of the propagator in replica space. The details of this will be explained in upcoming work. For instance, these these arguments can be used to show that the simplest diagram, involving just one $\delta G_\perp^q$, vanishes

$$N\frac{J^2}{2q}s_{ab}\int \mathrm{d}t\mathrm{d}t' \,\langle \delta G_\perp(t,t')_{ab}^q\rangle = 0, \quad (6.3.21)$$

where the angular brackets represent an average weighted by the quadratic action (6.3.20). As in SYK with one time point, the leading correction for large N comes from summing diagrams one $\delta G_\perp^q$ vertex and one δG^q vertex

$$\frac{N^2}{2}\frac{J^4}{4q^2}s_{ab}s_{cd}\int \mathrm{d}t_1\mathrm{d}t_2\mathrm{d}t_3\mathrm{d}t_4\,\langle \delta G_\perp(t_1,t_2)_{ab}^q\delta G_{cd}(t_3,t_4)^q\rangle. \quad (6.3.22)$$

This term contributes at order $1/N^{q-2}$ since it computes a Feynman diagram with two vertices and q propagators. As before, the only nonzero Wick contractions involve a contraction between $(\delta G_\perp)_{ab}$ and δG_{cd}, which can be computed using the free propagator

$$\langle \delta G_\perp(t_1,t_2)_{ab}\delta G_{cd}(t_3,t_4)\rangle = \pi^\perp_{ab,a'b'}\left[\mathsf{G}_{a'c}(t_1,t_3)\mathsf{G}_{b'd}(t_2,t_4) - \mathsf{G}_{a'd}(t_1,t_4)\mathsf{G}_{b'c}(t_2,t_3)\right], \quad (6.3.23)$$

where $\pi^\perp$ is the projector onto the orthocomplement of the subspace spanned by matrices of the form (6.2.24). Using the fact that $s_{ab}s_{cd} = s_{ad}s_{bc}$, we can rewrite the contribution (6.3.22) as

$$\frac{q!}{N^{q-2}}\frac{J^4}{4q^2}\pi^\perp_{ab,a'b'}s_{a'd}s_{b'd}\int \mathrm{d}t_1\mathrm{d}t_2\mathrm{d}t_3\mathrm{d}t_4\,\mathsf{G}_{a'd}(t_1,t_4)^q\mathsf{G}_{b'c}(t_2,t_3)^q. \quad (6.3.24)$$

By writing $\pi^\perp = 1 - \pi$, it's not too difficult to see that the first term leads to a contribution proportional to the square of

$$k \int \mathrm{d}t\mathrm{d}t' \left(\mathsf{G}^q_{LL} - \mathrm{i}^q \mathsf{G}^q_{LR} - \mathrm{i}^q \mathsf{G}^q_{RL} + \mathsf{G}^q_{RR}\right) =: k\frac{2TE}{J^2}, \tag{6.3.25}$$

whilst the second term is smaller by a factor of k due to the projection. Altogether, (6.3.22) becomes

$$k(k-1)\frac{q!}{N^{q-2}}\left(\frac{ET}{q}\right)^2. \tag{6.3.26}$$

This means that the total contribution from these saddle points and the fluctuations around them is

$$\frac{\langle |Z(\mathrm{i}T)|^{2k}\rangle}{\langle |Z(\mathrm{i}T)|^2\rangle^k} \supset \left[1 + k(k-1)\frac{q!}{N^{q-2}}\left(\frac{ET}{q}\right)^2 + \dots\right] \times \begin{cases} k! & \text{for } q = 0 \bmod 4, \\ (2k-1)!! & \text{for } q = 2 \bmod 4. \end{cases} \tag{6.3.27}$$

The leading correction has the same dependence on k and N as in SYK with one time point, however, there is a potentially nontrivial dependence on time T contained in the quantity ET. More precisely, E is a function of the variable β_{aux}, so we really we ought to replace

$$E \to \int_0^\infty \mathrm{d}\beta_{\text{aux}} \mu(\beta_{\text{aux}}) E(\beta_{\text{aux}})^2. \tag{6.3.28}$$

It would be interesting to understand how this quantity depends on T. We leave this for further work.

6.4 $q = 2$ SYK

For $q = 2$ the discrete S_{2k} replica symmetry of the problem is enhanced to a continuous symmetry. After integrating out the Gaussian fields G we can write the result as

$$\begin{aligned} \langle |Z(\mathrm{i}T)|^{2k}\rangle &= \int D\Sigma\, e^{-I[\Sigma]} \\ \frac{1}{N} I[\Sigma] &= -\frac{1}{2}\log\det(\partial_t - \Sigma) + \frac{1}{4J^2}\mathrm{Tr}\Sigma^2. \end{aligned} \tag{6.4.1}$$

In the infrared limit, the derivative operator ∂_t can self-consistently be dropped from the action, resulting in an effective theory with equation of motion $\Sigma^2 = J^2$, or, more explicitly

$$\int_0^T \mathrm{d}s\, \Sigma_{ac}(t,s)\Sigma_{cb}(s,t') = J^2\delta_{ab}\delta(t-t'). \tag{6.4.2}$$

This equation is invariant under the transformation

$$\Sigma(t,t') \to \Sigma^U(t,t') = \int_0^T \int_0^T \mathrm{d}s\mathrm{d}s'\, U(t,s)\Sigma(s,s')U(t',s')^{\text{tr}}, \tag{6.4.3}$$

with $U(t,t')$ a matrix valued distribution which is antiperiodic with period T in each argument and satisfies

$$\int_0^T \mathrm{d}s\, U(t,s)U(t',s)^{\mathrm{tr}} = \delta(t-t'), \tag{6.4.4}$$

where $U(t,t')^{\mathrm{tr}}$ is its transpose. It's useful to define U^{tr} to be the operator with matrix elements given by transposing both the replica indices and time labels of U, i.e. $(U^{\mathrm{tr}})(t,t') = U(t',t)^{\mathrm{tr}}$ where on the RHS tr is the ordinary transpose. The transformation can then be concisely written as

$$\Sigma \to \Sigma^U = U\Sigma U^{\mathrm{tr}}, \qquad UU^{\mathrm{tr}} = 1. \tag{6.4.5}$$

This corresponds to an infinite enhancement of the emergent diff(S^1) symmetry for each replica of the $q > 2$ model.[3] The symmetries of the single replica theory in the infrared were considered in [126] alongside its potential connection to an SL($N,\mathbb{R}$) BF theory. Away from the strict infrared limit of the theory, this symmetry is explicitly broken by the presence of the derivative operator ∂_t in the action. Indeed, the full equation of motion is

$$\partial_t\Sigma - \Sigma^2 = -J^2, \tag{6.4.6}$$

and so for U to be a symmetry of the full theory it must commute with the derivative operator

$$U^{\mathrm{tr}}\partial_t U = \partial_t. \tag{6.4.7}$$

It is simple to check that this can be solved by a $U(t,t')$ which only depends on the difference of times, i.e. $U(t,t') = U(t-t')$, since

$$\begin{aligned}\int_0^T \mathrm{d}s\, U(s-t)^{\mathrm{tr}}\partial_s U(s-t') &= -\partial_{t'}\int_0^T \mathrm{d}s\, U(s-t)^{\mathrm{tr}}U(s-t') \\ &= -\partial_{t'}\delta(t-t').\end{aligned} \tag{6.4.8}$$

In going to the second line we used the constraint $UU^{\mathrm{tr}} = 1$. Since U is antiperiodic it's Fourier series may be written as

$$U(t) = \frac{1}{\sqrt{T}}\sum_{n\in\mathbb{Z}} U(\omega_n)e^{-\mathrm{i}\omega_n t}, \qquad \omega_n = \frac{2\pi(n+\frac{1}{2})}{T}. \tag{6.4.9}$$

In order to preserve the reality condition on Σ, we require U to be real valued and so $U(\omega_n) = U(-\omega_n)^*$. It is then straightforward to verify that the constraint $UU^{\mathrm{tr}} = 1$ implies that the Fourier modes $U_n = U(\omega_n)$ satisfy $U_nU_n^\dagger = 1$, i.e. $U_n \in U(2k)$. The action of this symmetry on the Fourier modes of Σ is given by

$$\Sigma_{mn} \to (\Sigma^U)_{mn} = U_m\Sigma_{mn}U_n^{\mathrm{tr}}. \tag{6.4.10}$$

To summarise, the $U(1)^{2k}\times S_{2k}$ symmetry of the problem for $q = 2 \bmod 4$ is infinitely enhanced for the special case that $q = 2$ to $\prod_{n\geq 0} U(2k)$.

[3] Indeed, for the $q = 2$ model, there is a diff(S^1) subgroup given by setting $U(t,t') = \phi'(t)^\Delta\delta(\phi(t)-t')$ for $\phi \in \mathrm{diff}(S^1)$. The constraint $U^{\mathrm{tr}}U = 1$ then fixes $\Delta = \frac{1}{2}$.

6.4.1 Saddle points

To simplify the saddle point equation

$$\partial_t \Sigma - \Sigma^2 = -J^2, \tag{6.4.11}$$

we make an ansatz that Σ only depends on the difference of times, i.e. $\Sigma(t,t') = \Sigma(t-t')$. With this ansatz, configurations which are not invariant under this diagonal time translation symmetry only contribute as fluctuations in the one-loop determinant in the saddle point approximation. As such, it is useful to warm up to the full problem by studying a slightly simpler path integral defined similarly to (6.4.1) but where we only integrate over time translation invariant configurations. Writing the Fourier series of Σ as

$$\Sigma(t) = \sum_{n\in\mathbb{Z}} \Sigma(\omega_n) e^{-\mathrm{i}\omega_n t}, \qquad \omega_n = \frac{2\pi(n+\frac{1}{2})}{T}, \tag{6.4.12}$$

the path integral decomposes into an infinite product of decoupled hermitian matrix integrals

$$\langle |Z(\mathrm{i}T)|^{2k} \rangle = 2^{kN} \prod_{n=0}^{\infty} \left(\frac{N}{2\pi J^2}\right)^{2k^2} \int_{\mathbb{R}^{(2k)^2}} \mathrm{d}\Sigma_n \exp\left\{ N \mathrm{Tr}\log\left(1 + \frac{\Sigma_n}{\mathrm{i}\omega_n}\right) - \frac{N}{2J^2}\mathrm{Tr}\Sigma_n^2 \right\}, \tag{6.4.13}$$

where

$$\mathrm{d}\Sigma_n = \prod_a \mathrm{d}(\Sigma_n)_{aa} \prod_{a<b} \sqrt{2}\,\mathrm{dRe}\,(\Sigma_n)_{ab} \sqrt{2}\,\mathrm{dIm}\,(\Sigma_n)_{ab}, \tag{6.4.14}$$

is the standard measure on the space of $2k \times 2k$ hermitian matrices, which is a copy of $\mathbb{R}^{(2k)^2}$. That $\Sigma_n = \Sigma(\omega_n)$ is hermitian follows from imposing the constraint $\Sigma_{ab}(t,t') = -\Sigma_{ba}(t',t)$ as well as requiring $\Sigma_{ab}(t,t')$ to be purely imaginary, which is required for convergence of the path integral. The factor $2^{N/2}$ for each replica is the contribution of the free determinant $\prod_{n=0}^{\infty}(n+1/2)$ which, after zeta function regularisation, gives $\sqrt{2}$. Since $Z(\mathrm{i}T)$ can only depend on the dimensionful variables J and T through the combination JT, it's value at $J=0$ simply computes the dimension of the Hilbert space $Z(\mathrm{i}T)|_{J=0} = 2^{N/2}$. The normalisation chosen in (6.4.13) thus ensures that the spectral form factor is normalised on average.

Since the path integral (6.4.13) decomposes as product of decoupled matrix integral for each mode, we can focus our attention on the contribution of a single mode

$$Z_n = \left(\frac{N}{2\pi J^2}\right)^{2k^2} \int_{\mathbb{R}^{(2k)^2}} \mathrm{d}\Sigma_n \exp\left\{ N \mathrm{Tr}\log\left(1 + \frac{\Sigma_n}{\mathrm{i}\omega_n}\right) - \frac{N}{2J^2}\mathrm{Tr}\Sigma_n^2 \right\}. \tag{6.4.15}$$

The matrix integral over each mode Σ_n (6.4.13) is invariant under conjugation by a unitary matrix $\Sigma_n \to U_n \Sigma_n U_n^\dagger$ with $U_n \in U(2k)$. This symmetry is simply the restriction of the symmetry (6.4.10) mentioned in the previous section to time translation

invariant configurations. Using this symmetry to "gauge fix" each Σ_n to be a diagonal matrix gives

$$Z_n = \left(\frac{N}{2\pi J^2}\right)^{2k^2} \mathrm{vol}\left(\frac{U(2k)}{U(1)^{2k}\times S_{2k}}\right)\int_{\mathbb{R}^{2k}}\prod_a \mathrm{d}\lambda_a \prod_{a<b}(\lambda_a-\lambda_b)^2 e^{-N\sum_a V(\lambda_a)}, \quad (6.4.16)$$

where

$$V(\lambda) = -\log\left(1+\frac{\lambda}{\mathrm{i}\omega_n}\right)+\frac{1}{2J^2}\lambda^2. \quad (6.4.17)$$

With our normalisation

$$\mathrm{vol}\, U(x) = \frac{(2\pi)^{\frac{x(x+1)}{2}}}{G(x+1)}, \quad (6.4.18)$$

where G is the Barnes-G function. For large N, the integral (6.4.16) is dominated by it's saddle points and small fluctuations around them. For $k \ll N$, we can neglect the contribution of the Vandermonde determinant $\Delta(\{\lambda\}) = \prod_{a<b}|\lambda_a-\lambda_b|$ when determining the saddle points. The saddle points are then determined by extremising $V(\lambda)$ which gives a quadratic equation

$$\lambda_a^2 + \mathrm{i}\omega_n\lambda_a - J^2 = 0. \quad (6.4.19)$$

The two possible solutions for each eigenvalue are

$$\lambda_\pm = \frac{-\mathrm{i}\omega_n \pm \sqrt{4J^2-\omega_n^2}}{2}. \quad (6.4.20)$$

For $\omega_n < 2J$, the contour of integration for each eigenvalue may be deformed to pass through both saddle points $\lambda_\pm$, whilst for $\omega_n > 2J$ it is only possible to deform the contour to pass through the saddle point λ_+. First consider the simpler case that $\omega_n > 2J$. Expanding around the saddle point $\lambda_a = \lambda_+ + \mu_a$ we get

$$\begin{aligned} Z_n = e^{-2kNV(\lambda_+)}&\left(\frac{N}{2\pi J^2}\right)^{2k^2}\mathrm{vol}\left(\frac{U(2k)}{U(1)^{2k}\times S_{2k}}\right) \\ &\times\int_{\mathbb{R}^{2k}}\prod_a \mathrm{d}\mu_a\prod_{a<b}(\mu_a-\mu_b)^2 e^{-\frac{N}{2J^2}\left(1+\frac{\lambda_+^2}{J^2}\right)\sum_a\mu_a^2}. \end{aligned} \quad (6.4.21)$$

The one-loop contribution is most easily computed by interpreting the integral over μ_a as the integral over eigenvalues of a $2k\times 2k$ hermitian matrix X. We then have

$$\begin{aligned} Z_n &= e^{-2kNV(\lambda_+)}\left(\frac{N}{2\pi J^2}\right)^{2k^2}\int_{\mathbb{R}^{(2k)^2}}\mathrm{d}X\, e^{-\frac{N}{2J^2}\left(1+\frac{\lambda_+^2}{J^2}\right)\mathrm{Tr}X^2} \\ &= e^{-2kNV(\lambda_+)}\left(1+\frac{\lambda_+^2}{J^2}\right)^{-2k^2}. \end{aligned} \quad (6.4.22)$$

For $\omega_n < 2J$ each eigenvalue integral is given by a sum of over two saddle points. Since the integral is invariant under permutations of eigenvalues there are $2k+1$ distinct contributions

$$Z_n = \sum_{\ell=0}^{2k} Z_n^{(\ell)}, \tag{6.4.23}$$

where $Z_n^{(\ell)}$ is the contribution where ℓ eigenvalues are expanded around the λ_- saddle point and the remaining $2k-\ell$ eigenvalues are expanded around the λ_+ saddle point. Expanding around the saddle points $\lambda_a = \lambda_- + \mu_a^-$ for $a = 1, \ldots, \ell$ and $\lambda_a = \lambda_+ + \mu_a^+$ for $a = \ell+1, \ldots, 2k-\ell$ we get

$$\begin{aligned} Z_n^{(\ell)} = e^{-\ell N V(\lambda_-) - (2k-\ell) N V(\lambda_+)} \left(\frac{N}{2\pi J^2}\right)^{2k^2} (\lambda_+ - \lambda_-)^{2\ell(2k-\ell)} \mathrm{vol}\left(\frac{U(2k)}{U(1)^{2k} \times S_{2k}}\right) \\ \times \int_{\mathbb{R}^\ell} \prod_a \mathrm{d}\mu_a^- \prod_{a<b} (\mu_a^- - \mu_b^-)^2 e^{-\frac{N}{2J^2}\left(1+\frac{\lambda_-^2}{J^2}\right)\sum_a (\mu_a^-)^2} \\ \times \int_{\mathbb{R}^{2k-\ell}} \prod_a \mathrm{d}\mu_a^+ \prod_{a<b} (\mu_a^+ - \mu_b^+)^2 e^{-\frac{N}{2J^2}\left(1+\frac{\lambda_+^2}{J^2}\right)\sum_a (\mu_a^+)^2}. \end{aligned} \tag{6.4.24}$$

As before, the one-loop contributions are most easily computed by interpreting the integral over μ_a^- as the integral over eigenvalues of an $\ell \times \ell$ hermitian matrix and the integral over μ_a^+ as the integral over eigenvalues of a $(2k-\ell) \times (2k-\ell)$ hermitian matrix. This leads to the following result

$$\begin{aligned} Z_n = \sum_{\ell=0}^{2k} e^{-\ell N V(\lambda_-) - (2k-\ell) N V(\lambda_+)} \left(\frac{N}{2\pi J^2}(\lambda_+ - \lambda_-)^2\right)^{\ell(2k-\ell)} \mathrm{vol}\left(\frac{U(2k)}{U(\ell) \times U(2k-\ell)}\right) \\ \times \left(1 + \frac{\lambda_-^2}{J^2}\right)^{-\frac{\ell^2}{2}} \left(1 + \frac{\lambda_+^2}{J^2}\right)^{-\frac{(2k-\ell)^2}{2}}. \end{aligned} \tag{6.4.25}$$

This formula has a simple interpretation in terms of the original matrix integral (6.4.15). Due to the $U(2k)$ invariance of the matrix integral the extremums of $\mathrm{Tr} V(\Sigma_n)$ are highly degenerate. In general they are not saddle points, rather they are saddle manifolds. Let $\Lambda(\ell) = \mathrm{diag}(\lambda_-, \ldots, \lambda_-, \lambda_+, \ldots, \lambda_+)$, where ℓ denotes the multiplicity of λ_-. Then, each of the terms in sum above has the interpretation as the result of the integral along the orbits of the diagonal saddle points $\Lambda(\ell)$ under the action of $U(2k)$ together with the one-loop contribution in the directions orthogonal to the orbit. The orbit space is the quotient $U(2k)/G$ where G is the subgroup of $U(2k)$ which leaves the diagonal saddle point fixed. For $\Lambda(\ell)$, $G = U(\ell) \times U(2k-\ell)$ so the orbit space is the Grassmannian $\mathrm{Gr}(\ell, 2k) = U(2k)/U(\ell) \times U(2k-\ell)$, the space of ℓ-dimensional linear subspaces of $\mathbb{C}^{2k}$. The measure on the orbit comes from commuting $\mathfrak{g}^\perp$ with $\Lambda(\ell)$, where $\mathfrak{g}^\perp$ is the orthocomplement to the Lie algebra $\mathfrak{g}$ of G. Concretely, $\mathfrak{g}^\perp$ is the

$2\ell(2k-\ell)$-dimensional vector space spanned by $2k\times 2k$ antihermitian matrices that have vanishing entries in the top-left $\ell\times\ell$ block and bottom-right $(2k-\ell)\times(2k-\ell)$ block. The commutator of any matrix in $\mathfrak{g}^{\perp}$ with $\Lambda(\ell)$ is proportional to $\lambda_+-\lambda_-$, so the volume of the orbit is proportional to $(\lambda_+-\lambda_-)^{2\ell(2k-\ell)}$. This gives rise to the last two terms in the first line of (6.4.25).

Since $\mathrm{Re}V(\lambda_+)=\mathrm{Re}V(\lambda_-)$ the dominant contribution in the sum over saddles (6.4.25) is the term with $\ell=k$ as it maximises the volume of the orbit space, which scales as $N^{\ell(2k-\ell)}$. Only keeping this contribution gives

$$Z_n=e^{-kN[V(\lambda_-)+V(\lambda_+)]}\left(\frac{N}{2\pi J^2}(\lambda_+-\lambda_-)^2\right)^{k^2}\mathrm{vol}\left(\frac{U(2k)}{U(k)^2}\right)\left(1+\frac{\lambda_-^2}{J^2}\right)^{-\frac{k^2}{2}}\left(1+\frac{\lambda_+^2}{J^2}\right)^{-\frac{k^2}{2}}. \tag{6.4.26}$$

Using the formula (6.4.20) for $\lambda_\pm$ we get

$$Z_n=\left(\frac{J}{\omega_n}\right)^{2kN}e^{kN\left(\frac{\omega_n^2}{2J^2}-1\right)}\left(\frac{N}{2\pi}\sqrt{4-\frac{\omega_n^2}{J^2}}\right)^{k^2}\mathrm{vol}\left(\frac{U(2k)}{U(k)^2}\right). \tag{6.4.27}$$

Taking $k\gg 1$ but $k\ll N$ we find

$$Z_n\sim\left(\frac{J}{\omega_n}\right)^{2kN}e^{kN\left(\frac{\omega_n^2}{2J^2}-1\right)}\left(\frac{N}{k}\sqrt{4-\frac{\omega_n^2}{J^2}}\right)^{k^2}, \tag{6.4.28}$$

indicating a change of behaviour as k approaches a fixed fraction of N.

For early times such that $JT<\frac{\pi}{2}$, all the frequencies $\omega_n>2J$, so using (6.4.22) the spectral form factor is given by the infinite product

$$\langle|Z(\mathrm{i}T)|^{2k}\rangle=2^{kN}\prod_{n=0}^{\infty}e^{-2kNV(\lambda_+)}\left(1+\frac{\lambda_+^2}{J^2}\right)^{-2k^2}. \tag{6.4.29}$$

To leading order for small JT this gives

$$\langle|Z(\mathrm{i}T)|^{2k}\rangle=2^{kN}e^{-\frac{kN}{8}(JT)^2}. \tag{6.4.30}$$

For later times there will be a contribution from the frequencies ω_n with $\omega_n<2J$. Using (6.4.22) and (6.4.26), the spectral form factor is then given by the infinite product

$$\begin{aligned}\langle|Z(\mathrm{i}T)|^{2k}\rangle=2^{kN}\prod_{n=0}^{n_T}\left(\frac{J}{\omega_n}\right)^{2kN}e^{kN\left(\frac{\omega_n^2}{2J^2}-1\right)}\left(\frac{N}{2\pi}\sqrt{4-\frac{\omega_n^2}{J^2}}\right)^{k^2}\mathrm{vol}\left(\frac{U(2k)}{U(k)^2}\right)\\ \times\prod_{n=n_T+1}^{\infty}e^{-2kNV(\lambda_+)}\left(1+\frac{\lambda_+^2}{J^2}\right)^{-2k^2},\end{aligned} \tag{6.4.31}$$

where $n_T = \lfloor \frac{JT}{\pi} - \frac{1}{2} \rfloor$ is the number of frequencies ω_n with $\omega_n < 2J$. For large JT, the frequencies are very closely spaced so we can approximate the discrete variable ω_n by a continuous variable. This gives

$$\langle |Z(\mathrm{i}T)|^{2k} \rangle = 2^{kN} \left[\left(\frac{32N}{\pi e^3} \right)^{k^2} \mathrm{vol} \left(\frac{U(2k)}{U(k)^2} \right) \right]^{\frac{JT}{\pi}}. \tag{6.4.32}$$

The contribution of the classical action vanishes to leading order in JT as there is no exponential in N contribution (expect for the 2^{kN}), leaving only the contribution from the zero mode volume. Taking $k \gg 1$ but $k \ll N$ we find

$$\langle |Z(\mathrm{i}T)|^{2k} \rangle \sim 2^{kN} \left(\frac{16}{e^{3/2}} \frac{N}{k} \right)^{k^2 \frac{JT}{\pi}}. \tag{6.4.33}$$

This indicates a significant change in behaviour of the moments as k approaches a fixed fraction of N. In particular, naively extrapolating to $k \gg N$ suggests that the moments begin to decrease. We will return to this problem in section 6.4.2.

In these formulas we have only considered the one loop contribution from configurations which are invariant under a diagonal time translation. The full one loop contribution can be understood analytically for late times T. In this limit the dominant contribution is from the soft modes, which corresponds to making the zero mode configurations depend slightly on $t+t'$. The calculation is a straightforward generalisation of the one in [124], so we defer the details to appendix D.1. The result is

$$\langle |Z(\mathrm{i}T)|^{2k} \rangle = 2^{kN} \left[\left(\frac{8N}{\pi e^2 JT} \right)^{k^2} \mathrm{vol} \left(\frac{U(2k)}{U(k)^2} \right) \right]^{\frac{JT}{\pi}}. \tag{6.4.34}$$

The main effect is that N dependence has been rescaled, $N \to N/JT$, but still the that the moments grow very quickly with k in sharp contrast to what we found in the $q > 2$ model. Notably, this formula indicates a breakdown of the saddle point approximation as JT approaches a fixed fraction of N, which corresponds to the beginning of plateau region. Whilst we don't have an understanding of how to study the plateau using the collective field variables we can understand what the result should be based on a long-time average argument. By doing the Gaussian integral over the fermions, the unaveraged spectral form factor is

$$|Z(\mathrm{i}T)|^2 = 2^N \prod_{i=1}^{N/2} \cos^2 \frac{T\lambda_i}{2}, \tag{6.4.35}$$

where $\pm \mathrm{i}\lambda_i$ are the eigenvalues of J_{ij}. A long-time average implies the moments are

$$\langle |Z(\mathrm{i}T)|^{2k} \rangle = \binom{2k}{k}^{N/2}. \tag{6.4.36}$$

6.4.2 Large k from duality

For $k \gg N$ it is no longer possible to neglect the contribution of the Vandermonde determinant in (6.4.16) when determining the saddle points. However, there is a simple way to rewrite the original $2k \times 2k$ matrix integral (6.4.15)

$$Z_n = \left(\frac{N}{2\pi J^2}\right)^{2k^2} \int_{\mathbb{R}^{(2k)^2}} \mathrm{d}\Sigma e^{-\frac{N}{2J^2}\mathrm{Tr}\Sigma^2} \det\left(\mathrm{i}\omega_n + \Sigma\right)^N . \tag{6.4.37}$$

as an $N \times N$ matrix integral where the saddle point analysis is simple for $k \gg N$. For sake of notation we have added back the contribution from the free determinant in this expression. We will subtract it's contribution in the final expression. This duality is a special case of a duality found by Hikami and Brézin [127]. We start by introducing complex Grassmann variables χ_i^a and $\bar{\chi}_i^a$ with $i = 1, \ldots, N$ and $a = 1, \ldots, 2k$ to rewrite the determinant as

$$\det\left(\mathrm{i}\omega_n + \Sigma\right)^N = \int \mathrm{d}\chi \mathrm{d}\bar{\chi} e^{\mathrm{Tr}B(\mathrm{i}\omega_n + \Sigma)}, \tag{6.4.38}$$

where $\mathrm{d}\chi\mathrm{d}\bar{\chi} = \prod_{i,a} \mathrm{d}\chi_i^a \mathrm{d}\bar{\chi}_i^a$ and B is the $2k \times 2k$ matrix

$$B^{ab} = \sum_i \bar{\chi}_i^a \chi_i^b. \tag{6.4.39}$$

With the Grassmann integral representation of the determinant (6.4.38) the original matrix integral (6.4.37) becomes Gaussian. Integrating out Σ gives

$$Z_n = \int \mathrm{d}\chi \mathrm{d}\bar{\chi} e^{\mathrm{i}\omega_n \mathrm{Tr}B + \frac{J^2}{2N}\mathrm{Tr}B^2}. \tag{6.4.40}$$

By anticommuting the Grassmann variables we may write

$$\begin{aligned} \mathrm{Tr}B^2 &= \sum_{a,b}\sum_{i,j} \bar{\chi}_i^a \chi_i^b \bar{\chi}_j^b \chi_j^a \\ &= -\sum_{a,b}\sum_{i,j} \bar{\chi}_i^a \chi_j^a \bar{\chi}_j^b \chi_i^b \\ &= -\mathrm{Tr}\widetilde{B}^2, \end{aligned} \tag{6.4.41}$$

where $\widetilde{B}$ is the $N \times N$ matrix

$$\widetilde{B}_{ij} = \sum_a \bar{\chi}_i^a \chi_j^a. \tag{6.4.42}$$

Noting that also $\mathrm{Tr}B = \mathrm{Tr}\widetilde{B}$, we may rewrite (6.4.40) as

$$Z_n = \int \mathrm{d}\chi \mathrm{d}\bar{\chi} e^{\mathrm{i}\omega_n \mathrm{Tr}\widetilde{B} - \frac{J^2}{2N}\mathrm{Tr}\widetilde{B}^2}. \tag{6.4.43}$$

By introducing an $N \times N$ matrix $\widetilde{\Sigma}$ we may write

$$e^{-\frac{J^2}{2N}\mathrm{Tr}\widetilde{B}^2} = \left(\frac{N}{2\pi J^2}\right)^{\frac{N^2}{2}} \int_{\mathbb{R}^{N^2}} \mathrm{d}\widetilde{\Sigma} e^{-\frac{N}{2J^2}\mathrm{Tr}\widetilde{\Sigma}^2 + \mathrm{i}\mathrm{Tr}\widetilde{B}\widetilde{\Sigma}}. \tag{6.4.44}$$

Plugging this expression into (6.4.43) and integrating over the Grassmann variables gives

$$Z_n = \left(\frac{N}{2\pi J^2}\right)^{\frac{N^2}{2}} \int_{\mathbb{R}^{N^2}} \mathrm{d}\widetilde{\Sigma} e^{-\frac{N}{2J^2}\mathrm{Tr}\widetilde{\Sigma}^2} \det\left[\mathrm{i}(\omega_n + \widetilde{\Sigma})\right]^{2k}. \tag{6.4.45}$$

The duality exchanges the original integral over a $2k \times 2k$ hermitian matrix with a determinant insertion raised to the N^{th} power with an integral over an $N \times N$ hermitian matrix with a determinant insertion raised to the $2k^{\text{th}}$ power.[4] Finally, subtracting the contribution from the free determinant gives

$$Z_n = \left(\frac{N}{2\pi J^2}\right)^{\frac{N^2}{2}} \int_{\mathbb{R}^{N^2}} \mathrm{d}\widetilde{\Sigma} e^{-\frac{N}{2J^2}\mathrm{Tr}\widetilde{\Sigma}^2} \det\left(1 + \frac{\widetilde{\Sigma}}{\omega_n}\right)^{2k}. \tag{6.4.46}$$

The matrix integral over $\widetilde{\Sigma}$ is invariant under conjugation by a unitary matrix $\widetilde{\Sigma} \to U\widetilde{\Sigma}U^\dagger$ with $U \in U(N)$. Using this symmetry to gauge fix $\widetilde{\Sigma}$ to be a diagonal matrix gives

$$Z_n = \left(\frac{N}{2\pi J^2}\right)^{\frac{N^2}{2}} \mathrm{vol}\left(\frac{U(N)}{U(1)^N \times S_N}\right) \int_{\mathbb{R}^N} \prod_i \mathrm{d}\lambda_i \prod_{i<j} |\lambda_i - \lambda_j|^2 e^{-N\sum_i \widetilde{V}(\lambda_i)}, \tag{6.4.47}$$

where

$$\widetilde{V}(\lambda) = -\frac{2k}{N}\log\left(1 + \frac{\lambda}{\omega_n}\right) + \frac{1}{2J^2}\lambda^2. \tag{6.4.48}$$

For infinite k the saddle points for integral over eigenvalues are pushed to $\pm\infty$. For large but finite k the quadratic term in the potential provides a small stabilisation which pulls the saddle points from $\pm\infty$ to a finite distance.[5] The advantage of the dual description of the original matrix integral is that for $k \gg N$ we can self-consistently neglect the contribution from the Vandermonde determinant when determining the saddle points. The extremums of $\widetilde{V}(\lambda)$ are

$$\widetilde{\lambda}_\pm = \frac{-\omega_n \pm \sqrt{\frac{8k}{N}J^2 + \omega_n^2}}{2}. \tag{6.4.49}$$

[4] Performing this duality at the level of the full path integral basically undoes the G, Σ trick, see Appendix D.2.

[5] A similar mechanism it at play when determining the asymptotic behaviour of the gamma function using its integral representation.

The saddle point $\widetilde{\lambda}_+$ dominates for all ω_n and so

$$Z_n = e^{-N^2\widetilde{V}(\widetilde{\lambda}_+)}\left(1+\frac{N}{2k}\frac{\widetilde{\lambda}_+^2}{J^2}\right)^{-\frac{N^2}{2}}. \tag{6.4.50}$$

For large JT, the frequencies are very closely spaced so we can approximate the discrete variable ω_n by a continuous variable. This gives

$$\langle|Z(\mathrm{i}T)|^{2k}\rangle = 2^{kN}\exp\left[\frac{4}{3}\sqrt{2kN}\left(k-\frac{3(\pi-2)}{16}N\right)\frac{JT}{\pi}\right]. \tag{6.4.51}$$

This implies that the moments continue to grow as k increases beyond N, in contrast to the naive extrapolation of (6.4.33).

6.5 Sparse SYK

In section 6.3.4 we saw the first correction to $\langle|Z(\mathrm{i}T)|^{2k}\rangle/\langle|Z(\mathrm{i}T)|^2\rangle^k$ scales as k^2/N^{q-2}, so indicates a change in the behaviour of the large moments, where $k \sim N^{q/2-1}$. Interestingly, N^q is proportional to the number of independent random variables in the SYK model. In this section we try to understand in what sense this is the correct interpretation by considering a variant of the SYK model where we can tune the number of independent random variables.

The sparse SYK model [128] has gained a lot of attention recently because it reduces the computational complexity of the simulation. The sparse SYK model is defined as

$$H = \mathrm{i}^{q/2}\sum_{1\le i_1<\cdots<i_q\le N} J_{i_1\ldots i_q}x_{i_1\ldots i_q}\psi_{i_1}\ldots\psi_{i_q}, \tag{6.5.1}$$

where $x_{i_1\ldots i_q}$ is a Bernoulli random variable which takes the value 1 with probability p and the value 0 with probability $1-p$. The variance of the couplings are also slightly changed to account for the fact that many interactions are suppressed

$$\langle J_{i_1\ldots i_q}J_{j_1\ldots j_q}\rangle = \frac{1}{p}\frac{J^2(q-1)!}{N^{q-1}}\delta_{i_1j_1}\ldots\delta_{i_1j_q}. \tag{6.5.2}$$

It has been shown numerically that the SFF doesn't show significant deviation during the ramp regime so long as the sparsification probability p is greater than $1/N^{q-1}$.

In figure 6.2 we plot the higher moments of the SFF for various sparsification probabilities p, restricting to the case that p is larger than $1/N^{q-1}$ so that a linear ramp is still visible, even if it starts at later times. We see that the deviation is larger than RMT prediction, even for small values of k. In particular, the deviation increases as p decreases. This is inline with the expectation that the first correction to $\langle|Z(\mathrm{i}T)|^{2k}\rangle/\langle|Z(\mathrm{i}T)|^2\rangle^k$ is inversely related to the number of independent random variables.

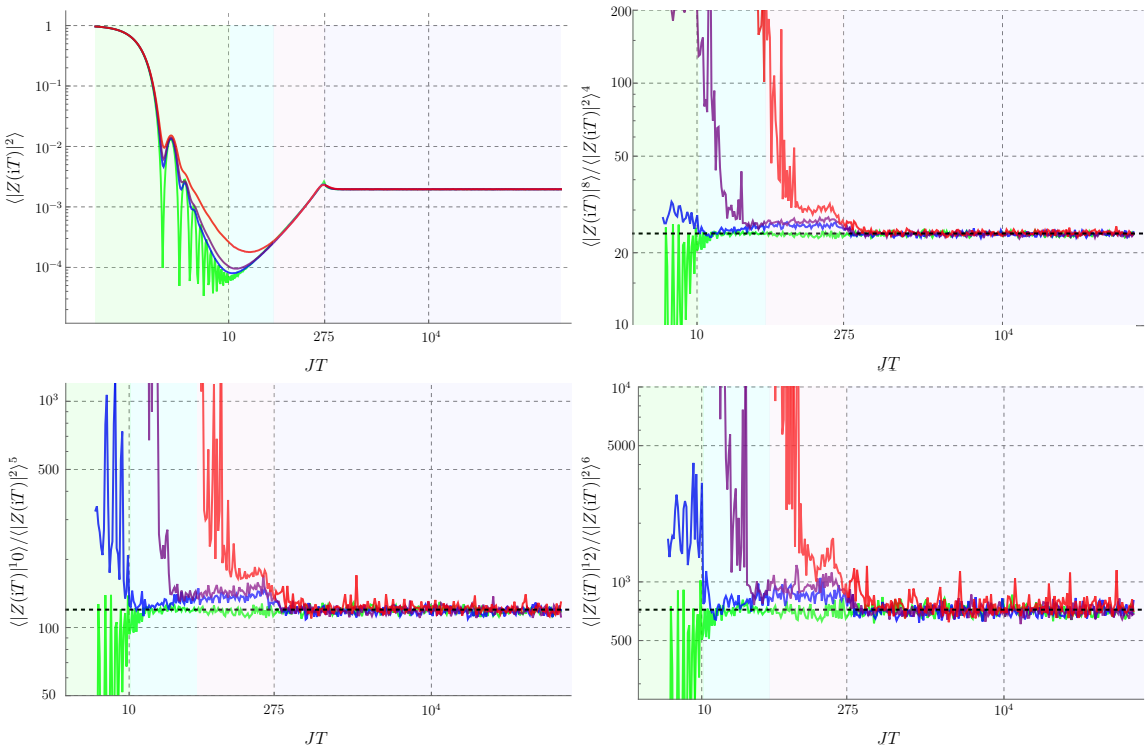


Figure 6.2: The $q = 4$ SYK model with $N = 20$ averaged over 2×10^5 samples with sparsification probability $p = 1$ (green), $p = 0.02$ (blue), $p = 0.015$ (purple), and $p = 0.01$ (red). The black dashed line is $k!$

6.6 Conclusion and outlook

For a chaotic system, the SFF is not a self averaging quantity at late times, instead it has erratic oscillations around its mean signal. In this chapter, we focused on studying the statistics of this noise during the ramp region in the SYK model. This was achieved by studying the moments $\langle |Z(\mathrm{i}T)|^{2k} \rangle$ of the SFF. In the large N limit this quantity can be studied using the collective field variables G and Σ. We found that the saddle points for $\langle |Z(\mathrm{i}T)|^{2k} \rangle$ could be simply constructed from the saddle points of $\langle |Z(\mathrm{i}T)|^2 \rangle$, which was studied by [34]. These saddle points broke the $2k$ independent time translation symmetries $U(1)^{2k}$ down to k diagonal time translation symmetries $U(1)^k$. This lead to a power law T^k ramp for the moments. In addition, these saddle points broke replica symmetry, for $q = 0 \bmod 4$: $S_{2k} \to S_k$. This latter fact was responsible for the non self averaging behaviour of the SFF.

Including $1/N$ corrections around the saddle point we found that this simple behaviour of the moments eventually breaks down, for sufficiently large k. More specifically, for $k \sim N^{q/2-1}$. By studying a sparse version of SYK in section 6.5 we saw numerically how this correction increases as we make the model more sparse. This

behaviour is inline with the intuition that the correction scales inversely with the number of random variables $\sim N^q$ for large N. It would be interesting to understand this behaviour analytically.

For $q = 2$ SYK we found that the behaviour of the moments is drastically different. In particular the fluctuations around the mean were much, much larger. This was related to the infinitely enlarged symmetry of the model, there was a $U(2k)$ symmetry associated to each Fourier mode of Σ. We were also able to understand the large k behaviour of the moments in this model using Hikami and Brézin duality.

We have studied the $q = 2$ and $q > 2$ models separately. It would also be interesting to study a model like $H_\lambda = \lambda H_{q=2} + (1-\lambda)H_{q>2}$, where $\lambda \in [0, 1]$. This Hamiltonian interpolates between the chaotic $q > 2$ model for $\lambda = 0$ and the integrable $q = 2$ model for $\lambda = 1$. It would be interesting to study the SFF and its moments in this model as a way to understand the transition from integrable and chaotic dynamics. Some aspects of this problem have been studied numerically in [129].

Another interesting problem would be to understand Maldacena's version of the information problem [40], as was briefly introduced in chapter 1, in the SYK model. More precisely, it would be interesting to understand the saddle point configurations for the collective fields variables which govern the two-point functions of operators like $\mathcal{O}(t) = \frac{1}{N}\sum_i \psi_i(t)\partial_t^n \psi_i(t)$.

Chapter 7

Conclusions

This thesis explored aspects of the black hole information problem, focusing primarily on the simple two-dimensional model of JT gravity reviewed in chapter 2. A central tool in this analysis was the QES prescription, a generalisation of the Ryu-Takayanagi proposal in the context of the AdS/CFT correspondence. Notably, this prescription turns out to be a signature of a deeper relation known as a EWR. Equipped with these ideas, this thesis explored aspects such as the behaviour of entropy in subsets of the Hawking radiation and information recovery from black holes, aiming to understand how the results compare with expectations based on random unitary models, like Page's model. These results provide nontrivial checks that, according to the QES prescription, black holes do indeed behave like a strongly chaotic quantum system with an effective Hilbert space dimension determined by the Bekenstein-Hawking entropy.

Below, we summarise the main findings of this thesis and propose some directions for future research.

In chapter 3 we saw how the entropy of disjoint subsets of the Hawking radiation is governed by a nontrivial competition between multiple QESs and captured by the *islands in the stream formula* (3.1.10). This is in sharp contrast to the Page curve calculation which involves the entropy of all the radiation and is governed by a competition between just two QESs. We studied this problem in a model of JT gravity coupled to a non-gravitational bath, examining cases with both a transparent and semi-transparent interface between the systems. The latter serves a toy model for a black hole with greybody factors. To understand the origin of multiple competing QESs, we studied a generalisation of Page's model where the dynamics of the black hole is modelled by a nested sequence of random isometries. This allowed for the process of evaporation to be thermodynamically irreversible. In this model we were able to precisely trace the origin of multiple QESs for disjoint subsets to irreversibility, recovering the case of only two QESs in the reversible limit.

Although these results enable one to study the correlation between subsets of the

radiation in the form of mutual information, they do not provide insight into the multipartite entanglement structure of the radiation. Exploring measures of multipartite entanglement, like entanglement negativity, which already exhibits a rich structure in random tensor networks [88], could provide valuable intuition in understanding this problem.

In chapter 4, we studied the Hayden-Preskill decoding criterion in detail in JT gravity coupled to a bath with conformal matter, by modelling the infalling system as a shockwave created by a local quench in the CFT. This model is rich enough to capture the near-horizon dynamics of the s-wave sector of a near-extremal black hole in higher dimensions, yet simple enough that the backreaction of the infalling system could be solved exactly. We explicitly verified the decoding criterion of Hayden and Preskill and its refinement, due to Hayden and Penington, for which understanding the effect of the backreaction was essential, and discovered additional effects caused by the irreversibility of evaporation and the backreaction of the infalling system.

From a holographic perspective, JT gravity coupled to a bath with conformal matter has a dual description as a quantum mechanical system coupled to a bath with the same conformal matter. In an appropriate infrared limit, BCFT offers a powerful framework to examine this setup. This motivated the work in chapter 5 where we considered two related models: an eternal black hole in JT gravity coupled to baths, and free fermion BCFT in the thermofield double state. At high temperatures, the dynamics of entropies in the BCFT are controlled by certain OPE limits of twist operator correlation functions used to compute the Rényi entropy. We demonstrated how this could be connected to the behaviour of certain OPE limits of twist operators in the bulk matter CFT, which can involve contributions from non-trivial QESs. This revealed how memory effects in the entropy of the bulk matter CFT impact the entropy of the radiation, an aspect not present in doubly holographic setups, where the bulk matter CFT is assumed to be holographic. It would be interesting to see how these memory effects are impacted by relaxing the limit of high temperatures and small scrambling times.

For chaotic quantum systems, the SFF is expected to exhibit noisy behaviour at late times. In systems described by ensembles, like SYK, ensemble averaging largely removes this noise. Nevertheless, one can study the statistics of these noisy fluctuations through the moments of the SFF. In chapter 6, we studied the moments of the SFF in SYK through the collective field description, which is closely related to the dual gravitational description. We saw how for large N the result agrees with expectations from RMT, whilst for any large but finite N, the connection breaks down for the very high moments of the SFF. We also studied the SFF in the $q = 2$ SYK model, an example of an integrable system. It would be interesting to extend our results to a system which interpolates between the $q = 2$ model and a $q > 2$ model, as a way to better understand the transition from integrability to chaos. Another obvious aspect to investigate would be the saddle points in the SYK model with fixed couplings and their interpretation in terms of half-wormholes [125].

Appendix A

Appendix for Chapter 2

A.1 K

The extrinsic curvature of a surface Σ is defined as

$$K_{ab} = h_a^c \nabla_c n_b, \tag{A.1.1}$$

where h_b^a is a projector onto tensors on Σ and n_a is the outward-pointing unit normal to Σ. In two dimensions and when Σ is a timelike curve we can write the projector as

$$h_b^a = -v^a v_b, \tag{A.1.2}$$

where v^a is the unit tangent to Σ. The trace of the extrinsic curvature may then be written as

$$K = v^a \nabla_v n_a \tag{A.1.3}$$

For the curve $t \mapsto (F(t), Z(t))$, written in Poincaré coordinates, the tangent v and outward-pointing unit normal n are

$$v = \frac{-Z(t)\left(F'(t)\partial_F + Z'(t)\partial_Z\right)}{\sqrt{F'(t)^2 - Z'(t)^2}}, \qquad n = \frac{-Z'(t)\mathrm{d}F + F'(t)\mathrm{d}Z}{-Z(t)\sqrt{F'(t)^2 - Z'(t)^2}}. \tag{A.1.4}$$

We assume the curve is future-directed so that $F'(t) > 0$. After a small calculation we find

$$K = \frac{F'(F'^2 - Z'^2 - ZZ'') + ZZ'F''}{(F'^2 - Z'^2)^{3/2}}, \tag{A.1.5}$$

where we have used that the nonvanishing Christoffel symbols are

$$\Gamma^F_{FZ} = \Gamma^F_{ZF} = \Gamma^Z_{FF} = \Gamma^Z_{ZZ} = -\frac{1}{Z}. \tag{A.1.6}$$

Using the relation $Z(t) = -\epsilon F'(t) + \mathcal{O}(\epsilon^3)$ and expanding for small ϵ gives

$$K = 1 - \epsilon^2\{F, t\} + \mathcal{O}(\epsilon^4). \tag{A.1.7}$$

Appendix B

Appendix for Chapter 3

B.1 Solving for the QESs

In this section we find a class of island that extremise the generalised entropy for a subset R of radiation with endpoints labelled by outgoing null coordinates u_i and ordered such that $u_1 > u_2 > \cdots > u_n$. For a given subset R we cannot claim to find all the possible islands which may contribute. For example, in appendix C of [93] an extremum is found which is not in the class considered here, however, this extremum is a maximum of the generalised entropy.

Consider a QES ∂I consisting of a collection of p points with Kruskal coordinates (U_a, V_a). Note that n and p must be even. The points ∂I and ∂R are ordered so that

$$U_a \gg U_b, \quad a > b, \qquad |U_i| \ll |U_j|, \quad i < j, \tag{B.1.1}$$

with $U_a > 0$ (that is, behind the horizon) and $U_u < 0$ (that is, outside the horizon). The condition above means that the subsets of radiation and spaces between them are large[1] compared with the thermal scale so that the thermodynamic formula for the entropy holds (3.1.6).

We now turn to the generalised entropy. We make the ansatz that the QES are close to the horizon in the sense that $U_a V_a \ll 1$. This will be verified ex post facto. The contribution from the QES can then be written using the universal near-horizon expression for the area term (2.5.23). The second term in the QES formula is the

[1] With respect to the measure $T(u)\mathrm{d}u$ or $\frac{\mathrm{d}U}{|U|}$.

entropy of quantum fields in the Unruh vacuum

$$\begin{aligned} S_{\mathrm{QFT}}(I \cup R) = &-\frac{N}{6}\sum_{a<b}(-1)^{a-b}\log\sigma_{ab} + \frac{N}{6}\sum_{a,i}(-1)^{a-i}\log\sigma_{ai} \\ &- \frac{N}{6}\sum_{i<j}(-1)^{i-j}\log\sigma_{ij} - \frac{N}{6}\sum_{a}\log\Omega_a - \frac{N}{6}\sum_{i}\log\Omega_i. \end{aligned} \tag{B.1.2}$$

Here $\sigma_{ab} = -(U_a - U_b)(v_a - v_b)$ etc, and the last two terms in (B.1.2), associated to the endpoints of ∂I and ∂R, are the contributions from the Weyl anomaly. For a point in ∂R this arises from rescaling the flat metric $-\mathrm{d}u\mathrm{d}v \to -\mathrm{d}U\mathrm{d}v$ which gives rise to the Weyl factor

$$\Omega_i^{-2} = \left.\frac{\mathrm{d}u}{\mathrm{d}U}\right|_{U=U_i} = -\frac{1}{2\pi T(u_i)U_i}. \tag{B.1.3}$$

For a point in ∂I, that is, a QES, rescaling the AdS_2 metric to $-\mathrm{d}U\mathrm{d}v$ gives rise to the Weyl factor

$$\Omega_a^{-2} = \frac{1}{(1+U_aV_a)^2}\left.\frac{\mathrm{d}V}{\mathrm{d}v}\right|_{V=V_a} \approx 2\pi T(v_a)V_a, \tag{B.1.4}$$

where we have assumed the near-horizon approximation $U_aV_a \ll 1$.

We first extremise the generalised entropy with respect to V_a. The terms $\log(v_a - v_b)$ will be subleading in the adiabatic limit so, henceforth, we drop their contribution. The negative flux of energy across the horizon (3.3.2) together with the first law imply

$$\mathrm{d}S_{\mathrm{BH}} = \frac{1}{T}\mathrm{d}E = -\frac{N}{24V}\mathrm{d}V. \tag{B.1.5}$$

Extremising the generalised entropy with respect to V_a then implies

$$-2(S_{\mathrm{BH}}(v_a) - S_0)U_a - \frac{N}{24V_a} + \frac{N}{12V_a} = 0. \tag{B.1.6}$$

Here the second term comes from (B.1.5) whilst the third term comes from the Weyl factor (B.1.4). This condition is precisely (3.3.4) and so $U_aV_a \ll 1$ as anticipated.

Extremising the generalised entropy with respect to U_a yields the coupled equations

$$2(S_{\mathrm{BH}}(v_a) - S_0)V_a + \frac{N}{6}\sum_{b(\neq a)}\frac{(-1)^{a-b}}{U_a - U_b} - \frac{N}{6}\sum_{j}\frac{(-1)^{a-j}}{U_a - U_j} = 0. \tag{B.1.7}$$

In order to solve the equations we can make use of the assumption (B.1.1) which says that the endpoints of ∂R are well separated with respect to the thermal scale in the sense that $|U_i| \ll |U_j|$ for $i < j$. To proceed we make an ansatz that to leading order

$$\tilde{u}_a = u_\alpha(a) + \ldots, \tag{B.1.8}$$

for a one-to-one map $\alpha : \partial I \to \partial R$ that preserves the order, that is, $\alpha(a) > \alpha(b)$ for $a > b$. This ansatz means that the reflection of the QESs ∂I across the horizon $\partial \tilde{I} \subset \partial R$. We now write the leading order correction as

$$U_a = -\lambda_a U_\alpha(a), \tag{B.1.9}$$

as solve for λ_a. The key insight is to notice the relative magnitudes of the outgoing Kruskal coordinates U of the QESs and ∂R:

$$\begin{array}{ccccccccc} |U_1| \ll \cdots \ll & |U_{\alpha(a)}| & \ll \cdots \ll & |U_{\alpha(b)}| & \ll \cdots \ll |U_n| \\ & \vdots & & \vdots & \\ \cdots \ll U_a \ll & \cdots & \ll U_b \ll \cdots & & \end{array} \tag{B.1.10}$$

for $a > b$, where the dotted arrows indicate terms of the same order. Using this, (B.1.7) becomes tractable

$$2(S_{\mathrm{BH}}(v_a) - S_0)V_a = \frac{N}{6}\begin{cases} (U_a - U_{\alpha(a)})^{-1} & a + \alpha(a) \in 2\mathbb{Z}, \\ U_a^{-1} - (U_a - U_{\alpha(a)})^{-1} & a + \alpha(a) \in 2\mathbb{Z} + 1. \end{cases} \tag{B.1.11}$$

Using (B.1.9) together with the near-horizon condition (3.3.4) it follows that

$$\lambda_a = \begin{cases} \frac{1}{3} & a + \alpha(a) \in 2\mathbb{Z}, \\ 3 & a + \alpha(a) \in 2\mathbb{Z} + 1. \end{cases} \tag{B.1.12}$$

This is the solution for the QES, including the leading order correction, which is precisely (3.3.8) if we identify the index i with $\alpha(a)$. Notice that at this order we can replace v_a in $T(v_a)$ and $S_{\mathrm{BH}}(v_a)$ with $u_{\alpha(a)}$.

We separate the contributions to the generalised entropy into four parts:

1. The contribution to $S_{\mathrm{QFT}}(I \cup R)$ is simply $S_{\mathrm{rad}}(R)$ in the thermodynamic limit.

2. The contribution to $S_{\mathrm{QFT}}(R \cup I)$ from the island is, to leading order,

$$\begin{aligned} S_{\mathrm{gen}}(I \cup R) \supset & -\frac{N}{6}\sum_{a<b}(-1)^{a-b}\log(U_b - U_a) + \frac{N}{12}\sum_a \log V_a \\ & -\frac{N}{6}\sum_{a<b}(-1)^{a-b}\log(U_{\alpha(a)} - U_{\alpha(b)}) - \frac{N}{12}\sum_a \log(-U_{\alpha(a)}) = S_{\mathrm{rad}}(\tilde{I}). \end{aligned} \tag{B.1.13}$$

   Notice that the infalling modes provides a contribution that looks like a Weyl factor because of the condition near-horizon condition (3.3.4), $\log U_a \sim -\log V_a$ at leading order.

3. The cross terms between the points in ∂I and points in ∂R are

$$S_{\text{gen}}(I \cup R) \supset \frac{N}{6} \sum_{a,j} (-1)^{a-j} \log(U_a - U_j) = -\frac{N}{6} \sum_{a,j} (-1)^{a-j} \log |U_{\max(\alpha(a),j)}| \\ = -2S_{\text{rad}}(\tilde{I} \cap R). \tag{B.1.14}$$

4. Finally the area term gives the contribution

$$S_{\text{gen}}(I \cup R) \supset \sum_{\substack{\text{disconnected} \\ \text{components} \\ \partial \tilde{I}_k \text{ of } \partial \tilde{I}}} S_{\text{BH}}(u_{\partial \tilde{I}_k}). \tag{B.1.15}$$

The result (3.1.7) then follows since the sum of the S_{QFT} contributions can be written as

$$S_{\text{QFT}}(I \cup R) \approx S_{\text{rad}}(\tilde{I}) + S_{\text{rad}}(R) - S_{\text{rad}}(\tilde{I} \cap R) = S_{\text{rad}}(\tilde{I} \ominus R). \tag{B.1.16}$$

B.2 Entropy flux inequality

We would like to show that

$$\frac{\mathrm{d}}{\mathrm{d}t} \left(S_{\text{rad}} - S^{\mathbb{R}}_{\text{rad}} + S_{\text{BH}} \right) = \frac{\pi N T}{6} \left(1 - \alpha^{\mathbb{R}} - \frac{\eta}{2} \right) \overset{?}{\geq} 0. \tag{B.2.1}$$

Using the explicit expressions for $\alpha^{\mathbb{R}}$ ((3.5.15) with $\Gamma \to 1 - \Gamma$) and η (3.5.13), we have

$$\alpha^{\mathbb{R}} + \frac{\eta}{2} = \frac{3}{\pi^2} \int_0^\infty \mathrm{d}x \left\{ \log[1 + \langle N^{\mathbb{R}}(x) \rangle] + \langle N^{\mathbb{R}}(x) \rangle \log \frac{1 + \langle N^{\mathbb{R}}(x) \rangle}{\langle N^{\mathbb{R}}(x) \rangle} + x \langle N^{\mathbb{T}}(x) \rangle \right\}. \tag{B.2.2}$$

For the first term in (B.2.2) we have

$$\log \left[1 + \langle N^{\mathbb{R}}(x) \rangle \right] \leq \left(1 + \frac{1}{e^x - 1} \right), \tag{B.2.3}$$

with equality if and only if $\Gamma = 0$. For the second term in (B.2.2) we have

$$\log \frac{\langle N^{\mathbb{R}}(x) \rangle}{1 + \langle N^{\mathbb{R}}(x) \rangle} \leq x, \tag{B.2.4}$$

with equality if and only if $\Gamma = 0$. Hence

$$\alpha^{\mathbb{R}} + \frac{\eta}{2} \leq \frac{1}{2} + \frac{3}{\pi^2} \int_0^\infty \mathrm{d}x \, x \left[\langle N^{\mathbb{R}}(x) + N^{\mathbb{T}}(x) \rangle \right] = 1. \tag{B.2.5}$$

Hence the bound (B.2.1) holds with equality if and only if $\Gamma = 0$.

B.3 Average Rényi entropy

In the random isometry model the n^{th} moment of ρ_R for K time steps is given by a sum over K copies of the symmetric group S_n (3.7.25). This motivates analysing the building blocks:

$$Z_n = \sum_{\pi \in S_n} d_1^{-d(\pi,\sigma_1)} d_2^{-d(\pi,\sigma_2)} d_3^{-d(\pi,\sigma_1)}, \tag{B.3.1}$$

where $d_k \geq 1$, $\sigma_k \in S_n$ and $d(\pi_1, \pi_2)$ is the Cayley distance function on S_n. When at least one of the parameters d_k is large, the dominant contributions to the sum are determined by minimising the "free energy"

$$F(\pi) = x_1 d(\pi, \sigma_1) + x_2 d(\pi, \sigma_2) + x_3 d(\pi, \sigma_3), \tag{B.3.2}$$

where $x_k = \log d_k$. We first consider the permutations which minimise the free energy at the following special regions in the *phase diagram* (see figure B.1), which we may parameterise by x_1/x_3 and x_2/x_3:

- for $x_1/x_3 \to 0$ and $x_2/x_3 \to 0$: $F(\pi) \to x_3 d(\pi, \sigma_3)$ is minimised for $\sigma = \sigma_3$.
- for $x_1/x_3 + x_2/x_3 = 1$: $F(\pi) = x_1 \left(d(\sigma_1, \pi) + d(\pi, \sigma_3)\right) + x_2 \left(d(\sigma_2, \pi) + d(\pi, \sigma_3)\right)$ is minimised for $\pi \in \Gamma(\sigma_1, \sigma_3) \cap \Gamma(\sigma_2, \sigma_3)$. Here $\Gamma(\sigma_k, \sigma_\ell)$ denotes the set of permutations π which saturate the triangle inequality $d(\sigma_k, \pi) + d(\pi, \sigma_\ell) \geq d(\sigma_k, \sigma_\ell)$.

There are four more regions in the phase diagram where the permutations which minimise the free energy can be determined by cyclically permuting the labels in the above. Most of the rest of the phase diagram can then be filled in using convexity of the free energy. That is, since F is a linear function of the x_k, if π minimises F at two points in the phase diagram, then π also minimises F along the segment joining these two points. This argument can only be used to fill in the whole phase diagram if the set of permutations $\Gamma(\sigma_1, \sigma_2, \sigma_3) \coloneqq \Gamma(\sigma_1, \sigma_2) \cap \Gamma(\sigma_2, \sigma_3) \cap \Gamma(\sigma_3, \sigma_1)$ which simultaneously saturate the three triangle inequalities

$$d(\sigma_k, \pi) + d(\pi, \sigma_\ell) \geq d(\sigma_k, \sigma_\ell) \qquad \text{for } k \neq \ell, \tag{B.3.3}$$

is not empty. The argument we have used to find the minima of F by considering special regions in the phase diagram and then using convexity to fill in the rest is due to [130]. This argument implies:

- for $x_1/x_3 + x_2/x_3 < 1$:

  $$Z_n \approx d_1^{-d(\sigma_1,\sigma_3)} d_2^{-d(\sigma_2,\sigma_3)}. \tag{B.3.4}$$

  The behaviour of the sum in two other regions may be obtained by cyclically permuting the labels in the above.

- assuming $\Gamma(\sigma_3, \sigma_2, \sigma_3)$ is not empty, for $x_1/x_3 + x_2/x_3 > 1$, $x_2/x_1 + x_3/x_1 > 1$ and $x_3/x_2 + x_1/x_2 > 1$:

$$Z_n \approx |\Gamma(\sigma_3, \sigma_2, \sigma_3)| \left(\frac{d_1 d_2}{d_3}\right)^{-d(\sigma_1,\sigma_2)/2} \left(\frac{d_2 d_3}{d_1}\right)^{-d(\sigma_2,\sigma_3)/2} \left(\frac{d_3 d_1}{d_2}\right)^{-d(\sigma_3,\sigma_1)/2}. \tag{B.3.5}$$

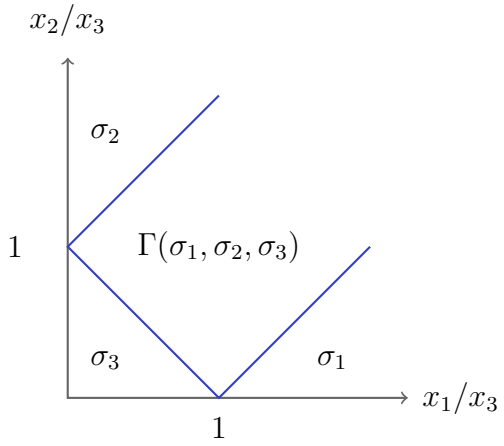


Figure B.1: Phase diagram for the sum (B.3.1) when $\Gamma(\sigma_1, \sigma_2, \sigma_3)$ is not empty. Along the blue lines there are more permutations which can contribute e.g. along $x_1/x_3 + x_2/x_3 = 1$ the sum is dominated by the set of permuations which lie in $\Gamma(\sigma_1, \sigma_3) \cap \Gamma(\sigma_2, \sigma_3)$.

We now turn to the proof that when the π_{N+1} and the σ_k are each either equal to the identity permutation id or the cyclic permutation τ, the sum

$$M(\pi_{K+1}) \coloneqq \mathbb{E}\mathrm{Tr}\rho_R^n \approx \prod_{k=1}^{K} \sum_{\pi_k \in S_n} d_{B_k}^{-d(\pi_k,\pi_{k+1})} d_{R_k}^{-d(\pi_k,\sigma_k)}, \tag{B.3.6}$$

is dominated by the terms where each of the π_k are themselves equal to the identity permutation id or the cyclic permutation τ. First, notice that $M(\pi_{K+1})$ satisfies the recursion relation

$$M(\pi_{K+1}) = \sum_{\pi_K \in S_n} d_{B_K}^{-d(\pi_K,\pi_{K+1})} d_{R_K}^{-d(\pi_K,\sigma_K)} M(\pi_K), \qquad M(\pi_1) = 1. \tag{B.3.7}$$

First consider

$$M(\pi_2) = \sum_{\pi_1 \in S_n} d_{B_1}^{-d(\pi_1,\pi_2)} d_{R_1}^{-d(\pi_1,\sigma_1)}. \tag{B.3.8}$$

This sum is of the form (B.3.1) so, by the above arguments, is dominated by the terms

with $\pi_1 \in \{\sigma_1, \pi_2\} \subset \{\mathrm{id}, \tau, \pi_2\}$. Using this fact we see that

$$\begin{aligned} M(\pi_3) &= \sum_{\pi_2 \in S_n} d_{B_2}^{-d(\pi_2,\pi_3)} d_{R_2}^{-d(\pi_2,\sigma_2)} M(\pi_2) \\ &\approx \sum_{\pi_2 \in S_n} d_{B_2}^{-d(\pi_2,\pi_3)} d_{R_2}^{-d(\pi_2,\sigma_2)} \min\{d_{B_1}, d_{R_1}\}^{-d(\pi_2,\sigma_1)}, \end{aligned} \tag{B.3.9}$$

is also of the form (B.3.1) so is dominated by the terms with $\pi_2 \in \{\sigma_1, \sigma_2, \pi_3\} \cup \Gamma(\sigma_1, \sigma_2, \pi_3) \subset \{\mathrm{id}, \tau, \pi_3\} \cup \Gamma(\mathrm{id}, \tau, \pi_3)$. Here we have assumed that $\Gamma(\mathrm{id}, \tau, \pi_3)$ is nonempty. Our strategy is to take this as an ansatz and then verify that this is consistent later. Using this, (B.3.4), and (B.3.5) it is simple to show that $M(\pi_4)$ is also of the form (B.3.1) so is dominated by the terms with $\pi_3 \in \{\mathrm{id}, \tau, \pi_4\} \cup \Gamma(\mathrm{id}, \tau, \pi_4)$. Again the strategy is to make the ansatz that $\Gamma(\mathrm{id}, \tau, \pi_4)$ is nonempty and then verify that this is consistent later. It is not too difficult to see that this pattern continues and proceeding with the argument we find that, provided $\Gamma(\mathrm{id}, \tau, \pi_{k+1})$ is not empty,

$$\tau_k \in \{\mathrm{id}, \tau, \pi_{k+1}\} \cup \Gamma(\mathrm{id}, \tau, \pi_{k+1}), \tag{B.3.10}$$

for each k. However, since $\tau_{K+1} \in \{\mathrm{id}, \tau\}$, by assumption, we have

$$\tau_k \in \{\mathrm{id}, \tau\}, \tag{B.3.11}$$

for each k. In particular, each $\Gamma(\mathrm{id}, \tau, \pi_{k+1})$ is nonempty, which is consistent with our assumption. Note that we have ignored the factors $|\Gamma(\dots)|$ which appear in (B.3.5). These terms depend on n but not on the dimensions of the subsystems d_{B_k} or d_{R_k} so do not play a role in determining which terms dominate the sum in the limit where the dimensions of the subsystems are large.

Appendix C

Appendix for Chapter 4

C.1 Solving for the QESs with a diary

In this section, we derive the generalisation of the island in the stream formula for a subset $R = \bigsqcup_k R_k$ of the Hawking radiation in the presence of a diary modelled as an infalling shockwave, as described in section 4.2. In this section we assume the matter sector consists of c massless Dirac fermions, as this is the only CFT for which the formula for the entropy of multiple intervals is known [79].

We label the QESs by $a, b, \ldots$, the endpoints of R by $i, j, \ldots$ and α to refer to both. The expression for the generalised entropy simplifies if we assume the QESs lie close to the apparent horizon. That is, the QESs lie close to $U = 0$ if $V_a < 1$ or close to $\tilde{U} = 0$, that is, the event horizon, if $V_a > 1$. More specifically,

$$U_a V_a \ll 1 \quad \text{for} \quad V_a < 1, \qquad \tilde{U}_a \tilde{V}_a \ll 1 \quad \text{for} \quad V_a > 1. \tag{C.1.1}$$

The strategy is to assume that this holds and then show that the solutions for the QESs are consistent with this assumption. In the adiabatic approximation, and assuming that the QESs lie close to the appropriate horizon, the area term contributes to the generalised entropy as

$$\frac{A}{4G} \sum_{a:V_a<1} S_{\mathrm{BH}}(v_a)(1 - 2U_a V_a) + \sum_{a:V_a>1} S_{\mathrm{BH}}(v_a)(1 - 2\tilde{U}_a \tilde{V}_a). \tag{C.1.2}$$

The QFT entropy contributes to the generalised entropy as

$$S_{\mathrm{QFT}} = -\frac{c}{6} \sum_{\alpha<\beta} (-1)^{\alpha-\beta} \log(U_\alpha - U_\beta) - \frac{c}{12} \sum_\alpha \log U'_\alpha + \frac{c}{12} \sum_a \log \frac{V'_a U'_a}{(1 + U_a V_a)^2}. \tag{C.1.3}$$

Here $V'_a = V'(v_a)$ and $U'_\alpha = U'(u_\alpha)$, where the coordinate u_a (what we call $\tilde{u}_a$ in the main text) is defined for the QES by $U_a = -U(u_a)$. In these expression we are viewing

U and V as the functions $t \mapsto -1/F(t)$ and $t \mapsto F(t)$, respectively. The above can also be expressed in terms of the $\tilde{U}$ and $\tilde{V}$ coordinates in an identical way. Moreover, each term in the third sum is invariant under a Möbius transformation. We now apply the assumption (C.1.1) that the QESs lie close to the apparent horizon to simplify (C.1.4). We find

$$\begin{aligned} S_{\text{QFT}} = &-\frac{c}{6}\sum_{\alpha<\beta}(-1)^{\alpha-\beta}\log(U_\alpha - U_\beta) - \frac{c}{12}\sum_{\alpha:|U_\alpha|>1}\log|U_\alpha| \\ &- \frac{c}{12}\sum_{\alpha:|U_\alpha|<1}\log|\tilde{U}_\alpha| + \frac{c}{12}\sum_{a:V_a<1}\log V_a U_a + \frac{c}{12}\sum_{a:V_a>1}\log\tilde{V}_a\tilde{U}_a. \end{aligned} \tag{C.1.4}$$

To obtain this expression it is useful to use the following formulas:

$$\frac{V'}{V}(t) = -\frac{U'}{U}(t) = 2\pi T(t) \quad \text{for } t < t_D, \qquad \frac{\tilde{V}'}{\tilde{V}}(t) = -\frac{\tilde{U}'}{\tilde{U}}(t) = 2\pi T(t) \quad \text{for } t > t_D. \tag{C.1.5}$$

In obtaining (C.1.4) we have dropped subleading terms of order $c\log T$ and have used the approximation $U = \tilde{U} - \lambda$ so that $U' \approx \tilde{U}'$, which is valid when $|U| < 1$.

Varying the generalised entropy with respect to V_a for a QES with $V_a < 1$ and $\tilde{V}_a$ for a QES with $V_a > 1$, we find

$$V_a U_a = \frac{c}{48 S_{\text{BH}}(v_a)} \ll 1, \qquad \tilde{V}_a\tilde{U}_a = \frac{c}{48 S_{\text{BH}}(v_a)} \ll 1. \tag{C.1.6}$$

Since $S_{\text{BH}} \gg c$ in the adiabatic approximation, this is consistent with our earlier assumption (C.1.1). These equations have the approximate solution

$$v_a = u_a - \frac{1}{2\pi T(u_a)}\log\frac{S_{\text{BH}}(u_a)}{c}, \tag{C.1.7}$$

where the second term is assumed to be small compared with the size of the intervals which means for any interval $R_k \subset R$, we have $S_{\text{rad}}(R_k) \gg c\log S_{\text{BH}}/c$.

We now turn to extremising the generalised entropy with respect to U_a. It's useful to consider the case that the QES lies to the past of D, that is, $v_a < t_D$, into two cases, depending whether $u_a \lessgtr t_D$. First consider the case that $u_a < t_D$. Extremising the generalised entropy with respect to U_a and using (C.1.6) we get

$$\begin{matrix} v_a < t_0 \\ u_a \lesssim t_0 \end{matrix} : \qquad \frac{1}{4} + \sum_{\alpha\neq a}(-1)^{a-\alpha}\frac{U_a}{U_a - U_\alpha} = 0, \tag{C.1.8}$$

which is familiar form the analysis in appendix B.1.

Solutions to this equation are simple because there is a hierarchy of scales amongst the coordinates due to the fact that endpoint of R are well separated compared to the

thermal scale. That is, if $u_i > u_j$, then $|U_i| \ll |U_j|$.[1] This implies that each QES lies close to a point in ∂R, in the sense that $U_a = \kappa |U_j|$ for some $\kappa = \mathcal{O}(1)$. The sum is then well approximated by

$$\begin{matrix} v_a < t_0 \\ u_a \lesssim t_0 \end{matrix} : \qquad \frac{1}{4} - \frac{U_a}{U_a - U_j} = 0 \qquad \Longrightarrow \qquad U_a = \frac{1}{3}|U_j|. \tag{C.1.9}$$

There is a slight subtlety here. If we order the coordinates $\{U_\alpha\} = \{U_j\} \cup \{U_a\}$ in terms of their magnitude $|U_\alpha|$, then the number of coordinates that have $|U_\alpha| \ll |U_j|$ must be even. There are solutions for the case that the number is odd, however these correspond to maxima of the generalised entropy and so don't play any role as far as the entropy is concerned.

Consider the case that $v_a < t_D$ but now with $u_a > t_D$. We still have (C.1.8), but the presence of the diary complicates the discussion since all the points in ∂R with $u_i > t_0$ have coordinates $U_i \sim -\lambda$ and we expect $U_a = \mathcal{O}(\lambda)$. We will now consider an ansatz of the form $U_a = \kappa\lambda$. In this case we can neglect all the terms in (C.1.8) with $|U_i| \gg \lambda$. All the other terms have an identical contribution. If we consider the case that there is an even number of coordinates with $\lambda \sim |U_\alpha| \ll 1$, then we find that all these terms exactly cancel out, leaving us with

$$\begin{matrix} v_a < t_D \\ \hat{u}_a > t_D \end{matrix} : \qquad \frac{1}{4} - \frac{U_a}{U_a + \lambda} = 0 \qquad \Longrightarrow \qquad U_a = \frac{1}{3}\lambda. \tag{C.1.10}$$

Finally, we consider the case that the QES lies to the future of the diary, that is, $v_a > t_D$. Extremising the generalised entropy with respect to U_a leads to:

$$v_a > t_D : \qquad \frac{1}{4} + \sum_{\alpha \neq a} (-1)^{a-\alpha} \frac{\tilde{U}_a}{U_a - U_\alpha} = 0. \tag{C.1.11}$$

Again, the presence of diary affects the discussion. However, the key point is that it does not change the hierarchy of scales amongst the coordinates, in the sense that, for $u_i > u_j$,

$$U_i - U_j \approx \begin{cases} -U_j & u_j < t_D, \\ -\tilde{U}_j & u_j > t_D. \end{cases} \tag{C.1.12}$$

Here we have used $U_j = -\lambda + \tilde{U}_j$ for $u_j > t_D$. So, the hierarchy of scales amongst the U_α coordinates can still be exploited to give a solution

$$v_a > t_D : \qquad \tilde{U}_a = \frac{1}{3}|\tilde{U}_j|. \tag{C.1.13}$$

[1] Notice that since the diary is assumed to be small, in the sense that $\lambda \ll 1$, the hierarchy between the Kruskal coordinates also holds when $u_a \sim t_D$, as indicated in equation (C.1.8), that is, until $|\tilde{U}_a| \gg \lambda$.

It is noteworthy that when $u_j > t_D$ there are two QESs that have $v_a \gtrless t_D$. When the QES lies to the past of D, we have $U_a = \lambda/3$ and, approximately,

$$v_a = t_D - \frac{1}{2\pi T(t_D)} \log \frac{\Delta S_D}{24c}. \tag{C.1.14}$$

The other solution has

$$v_a = u_j - \frac{1}{2\pi T(u_j)} \log \frac{S_{\rm BH}(u_j)}{c} > t_D. \tag{C.1.15}$$

Using (C.1.14) and (C.1.15), we can also compute the logarithmic corrections to the generalised entropy. Each QES gives rise to a logarithmic correction correction which gives

$$S_{\rm gen} = \sum_{\substack{\text{disconnected} \\ \text{components} \\ \partial\tilde{I}_k \text{ of } \partial\tilde{I}}} \left(S_{\rm BH}(u_{\partial\tilde{I}_k}) - \frac{c}{24} \log \frac{S_{\rm BH}(u_{\partial\tilde{I}_k})}{c} \right) + S_{\rm rad}(\tilde{I} \ominus R). \tag{C.1.16}$$

Actually for the case that one of the QESs is the one that gets stuck to the past of the diary we have to replace one of the logarithmic corrections

$$-\frac{c}{24} \log \frac{S_{\rm BH}(u_{\partial\tilde{I}_k})}{c} \to 3 \log \frac{\Delta S_{\rm BH}}{S_{\rm BH}(t_D)}. \tag{C.1.17}$$

This is a refinement of the islands in the stream formula (4.2.6).

Appendix D

Appendix for Chapter 6

D.1 One loop contribution for $q = 2$

In section 6.4 we only considered the one loop contribution from configurations which are invariant under a diagonal time translation. In this section we consider the full one loop contribution. Expanding in fluctuations around the saddle points, $\Sigma = \mathbf{\Sigma} + \delta\Sigma$, the quadratic action is

$$\delta I = \frac{N}{4J^2} \sum_{a,b} \int \mathrm{d}t_1 \dots \mathrm{d}t_4 \delta\Sigma_{ab}(t_1, t_2) K_{ab}(t_1, \dots, t_4) \delta\Sigma_{ba}(t_4, t_3), \tag{D.1.1}$$

where

$$K_{ab}(t_1, t_2, t_3, t_4) = K_{ab}(t_1 - t_3, t_2 - t_4) = \delta(t_1 - t_3)\delta(t_2 - t_4) + J^{-2}\mathbf{\Sigma}_{aa}(t_3 - t_1)\mathbf{\Sigma}_{bb}(t_2 - t_4). \tag{D.1.2}$$

In Fourier space, this reads

$$\delta I = \frac{N}{4J^2T^2} \sum_{a,b} \sum_{\omega_1,\omega_2} |\delta\Sigma_{ab}(\omega_1, \omega_2)|^2 K_{ab}(-\omega_2, -\omega_1), \tag{D.1.3}$$

where

$$K_{ab}(-\omega_2, -\omega_1) = 1 + J^{-2}\mathbf{\Sigma}_{aa}(\omega_1)\mathbf{\Sigma}_{bb}(-\omega_2). \tag{D.1.4}$$

In order to regularise the one loop determinant, we impose that the determinant goes to 1 as $J \to 0$. In this limit, $\mathbf{\Sigma} \sim J^2/\omega$ which leads to the following normalisation

$$\frac{\det\left(\frac{N}{4J^2T^2}K\right)^{-1/2}}{\det\left(\frac{N}{4J^2T^2}\right)^{-1/2}} = \exp\left[-\frac{1}{2}\sum_{a,b}\sum_{\omega_1,\omega_2} \log K_{ab}(\omega_2, \omega_1)\right] =: Z_{\text{one loop}}. \tag{D.1.5}$$

The kernel $K_{ab}(\omega_2, -\omega_1)$ is zero in the range $-2J < \omega_1 = \omega_2 < 2J$ with a, b indexing the off-diagonal blocks. These are the zero modes, whose contribution has be accounted for

in the previous section. In the previous section only configurations under the diagonal time translation symmetry were considered. Soft modes correspond to moving a little bit away from these zero modes, that is we consider $\omega_2 = \omega$, $\omega_1 = \omega + \epsilon$ with $\epsilon = 2\pi m/T$, where m is an integer and $|\epsilon| \ll 2J$. Notice that this only makes sense at late times. In the off-diagonal blocks we have

$$K_{ab}(\omega, -\omega - \epsilon) = \mathrm{i}\frac{\epsilon}{\sqrt{4J^2 - \omega^2}}, \tag{D.1.6}$$

so we can rewrite the one loop contribution as

$$\log Z_{\text{one loop}} = \frac{1}{2} \cdot 4 \cdot k^2 \sum_{m=1}^{M} \sum_{0<\omega<2J} \log\left(\frac{2\pi m}{T} \frac{1}{\sqrt{4J^2 - \omega^2}}\right). \tag{D.1.7}$$

where the factor of 4 comes from having restricted the sums on m, ω to just the positive values, while the sum over a, b gave the factor k^2. M is a positive integer counting the number of soft modes, $M \ll JT$. The presence of M is quite unpleasing since there is not a natural value for it. We can proceed as in [124] by using the analytical properties of the kernel $\tilde{K}$ and its behaviour at infinity. In particular, we have

$$\int \mathrm{d}\omega_1 \mathrm{d}\omega_2 \log K_{ab}(\omega_1, \omega_2) = 0. \tag{D.1.8}$$

The idea now is to subtract this integral to regulate the sum as

$$\log Z_{\text{one loop}} = 2k^2 \sum_{0<\omega<2J} \left(\sum_{m=1}^{M} - \int_0^{M+\frac{1}{2}} \mathrm{d}m\right) \log\left(\frac{2\pi m}{T} \frac{1}{\sqrt{4J^2 - \omega^2}}\right). \tag{D.1.9}$$

Using Stirling's approximation, we have

$$\begin{aligned}\left(\sum_{m=1}^{M} - \int_0^{M+\frac{1}{2}} \mathrm{d}m\right) \log(ma) &= \log(M! a^M) - \left(M + \frac{1}{2}\right)\left(\log\left(a\left(M + \frac{1}{2}\right)\right) - 1\right) \\ &= -\frac{1}{2}\log\frac{a}{2\pi},\end{aligned} \tag{D.1.10}$$

so we can write

$$\log Z_{\text{one loop}} = \sum_{0<\omega<2J} k^2 \log\left(T\sqrt{4J^2 - \omega^2}\right). \tag{D.1.11}$$

The next step is to perform the sum over $\omega \sim 2\pi n/T$

$$\begin{aligned}\log Z_{\text{one loop}} =& k^2 \frac{1}{2} \sum_{n<JT/\pi} \log\left[(2\pi)^2\left(\frac{J^2T^2}{\pi^2} - n^2\right)\right] = k^2 \log\left[(2\pi)^{\frac{2JT}{\pi}}\left(\frac{2JT}{\pi}\right)!\right] \\ =& -k^2 \frac{JT}{\pi} \log\left(\frac{e}{4JT}\right).\end{aligned} \tag{D.1.12}$$

Notice that for $k = 1$ this result matches [124, eq. (31)]. The ramp region, corrected with the soft modes contribution therefore reads:

$$\langle |Z(\mathrm{i}T)|^{2k} \rangle = 2^{kN} \left[\left(\frac{8N}{\pi e^2 JT} \right)^{k^2} \mathrm{vol} \left(\frac{U(2k)}{U(k)^2} \right) \right]^{\frac{JT}{\pi}}. \tag{D.1.13}$$

D.2 More on $q = 2$

A similar trick to the one used in section 6.4.2 can be applied to the full path integral (6.4.1)

$$\langle |Z(\mathrm{i}T)|^{2k} \rangle = \int D\Sigma e^{-\frac{N}{4J^2}\mathrm{Tr}\Sigma^2} \det(\partial_t - \Sigma)^{\frac{N}{2}}. \tag{D.2.1}$$

The analogue of the matrix B is

$$G_{ab}(t, t') = \frac{1}{N} \sum_i \psi_i^a(t) \psi_i^b(t'), \tag{D.2.2}$$

whilst the analogue of $\widetilde{B}$ is the antisymmetric matrix

$$\widetilde{G}_{ij} = \sum_a \int_0^T \mathrm{d}t \psi_i^a(t) \psi_j^a(t). \tag{D.2.3}$$

Going through the steps basically undoes the G, Σ trick and gives back

$$\langle |Z(\mathrm{i}T)|^{2k} \rangle = \left(\frac{N}{2\pi J^2} \right)^{\frac{N(N-1)}{4}} \prod_{i<j} \int \mathrm{d}J_{ij} e^{-\frac{N}{2J^2}J_{ij}^2} \int D\psi \exp\left\{ -\frac{1}{2} \int_0^T \mathrm{d}t \left(\psi_i^a \partial_t \psi_i^a - J_{ij} \psi_i^a \psi_j^a \right) \right\}. \tag{D.2.4}$$

In this expression there is an implicit sum over both i and a in the action. Of course, since the $q = 2$ model is quadratic in the fermions we could have integrated them out in the first step. This is simplest if we work in Fourier space. After decomposing the fermions into Fourier modes

$$\psi_i^a(t) = \frac{1}{\sqrt{T}} \sum_{n \in \mathbb{Z}} \psi_i^a(\omega_n) e^{-\mathrm{i}\omega_n t}, \qquad \omega_n = \frac{2\pi(n + \frac{1}{2})}{T}. \tag{D.2.5}$$

we get an infinite product of decoupled Grassmann integrals for each mode. This gives

$$\prod_{n=0}^{\infty} \det(\mathrm{i}\omega_n \delta_{ij} + J_{ij})^{2k} = 2^{kN} \prod_{i=1}^{N/2} \cos^{2k} \frac{T\lambda_i}{2}. \tag{D.2.6}$$

In this expression 2^{kN} is the contribution of the free determinant. By an orthogonal transformation, the antisymmetric matrix J_{ij} can be brought into block diagonal form with blocks

$$\begin{pmatrix} 0 & \lambda_i \\ -\lambda_i & 0 \end{pmatrix}, \qquad \lambda_i \geq 0. \tag{D.2.7}$$

Expressing the determinant in terms of the λ_i, $i = 1, \ldots, N/2$, and doing the infinite product over the modes give the second term in (D.2.6). Writing the matrix integral over J_{ij} as an integral over the λ_i gives

$$\langle |Z(\mathrm{i}T)|^{2k} \rangle = 2^{kN} \left(\frac{N}{2\pi J^2} \right)^{\frac{N(N-1)}{4}} \mathrm{vol} \left(\frac{O(N)}{SO(2)^{\frac{N}{2}} \times S_{\frac{N}{2}}} \right) \prod_i \int_0^\infty \mathrm{d}\lambda_i \prod_{i<j} (\lambda_i^2 - \lambda_j^2)^2 e^{-N \sum_i V(\lambda_i)}. \tag{D.2.8}$$

where

$$V(\lambda) = \frac{1}{2J^2} \lambda^2 - \frac{k}{N} \log \cos^2 \frac{T\lambda}{2}. \tag{D.2.9}$$

Bibliography

[1] Roger Penrose. "Gravitational collapse and space-time singularities". In: *Phys. Rev. Lett.* 14 (1965), pp. 57–59. DOI: 10.1103/PhysRevLett.14.57.

[2] J. D. Bekenstein. "Black holes and the second law". In: *Lett. Nuovo Cim.* 4 (1972), pp. 737–740. DOI: 10.1007/BF02757029.

[3] S. W. Hawking. "Particle Creation by Black Holes". In: *Commun. Math. Phys.* 43 (1975). Ed. by G. W. Gibbons and S. W. Hawking. [Erratum: Commun.Math.Phys. 46, 206 (1976)], pp. 199–220. DOI: 10.1007/BF02345020.

[4] S. W. Hawking. "Breakdown of Predictability in Gravitational Collapse". In: *Phys. Rev. D* 14 (1976), pp. 2460–2473. DOI: 10.1103/PhysRevD.14.2460.

[5] Don N. Page. "Information in black hole radiation". In: *Phys. Rev. Lett.* 71 (1993), pp. 3743–3746. DOI: 10.1103/PhysRevLett.71.3743. arXiv: hep-th/9306083.

[6] Samir D Mathur. "The information paradox: a pedagogical introduction". In: *Classical and Quantum Gravity* 26.22 (Oct. 2009), p. 224001. ISSN: 1361-6382. DOI: 10.1088/0264-9381/26/22/224001. URL: http://dx.doi.org/10.1088/0264-9381/26/22/224001.

[7] Joseph Polchinski. *String Theory and Black Hole Complementarity*. 1995. arXiv: hep-th/9507094 [hep-th]. URL: https://arxiv.org/abs/hep-th/9507094.

[8] Ahmed Almheiri, Thomas Hartman, Juan Maldacena, Edgar Shaghoulian, and Amirhossein Tajdini. "The entropy of Hawking radiation". In: *Reviews of Modern Physics* 93.3 (July 2021). ISSN: 1539-0756. DOI: 10.1103/revmodphys.93.035002. URL: http://dx.doi.org/10.1103/RevModPhys.93.035002.

[9] Ahmed Almheiri, Netta Engelhardt, Donald Marolf, and Henry Maxfield. "The entropy of bulk quantum fields and the entanglement wedge of an evaporating black hole". In: *JHEP* 12 (2019), p. 063. DOI: 10.1007/JHEP12(2019)063. arXiv: 1905.08762 [hep-th].

[10] Geoffrey Penington. "Entanglement Wedge Reconstruction and the Information Paradox". In: *JHEP* 09 (2020), p. 002. DOI: 10.1007/JHEP09(2020)002. arXiv: 1905.08255 [hep-th].

[11] Geoff Penington, Stephen H. Shenker, Douglas Stanford, and Zhenbin Yang. "Replica wormholes and the black hole interior". In: *JHEP* 03 (2022), p. 205. DOI: 10.1007/JHEP03(2022)205. arXiv: 1911.11977 [hep-th].

[12] Ahmed Almheiri, Thomas Hartman, Juan Maldacena, Edgar Shaghoulian, and Amirhossein Tajdini. "Replica Wormholes and the Entropy of Hawking Radiation". In: *JHEP* 05 (2020), p. 013. DOI: 10.1007/JHEP05(2020)013. arXiv: 1911.12333 [hep-th].

[13] G. W. Gibbons and S. W. Hawking. "Action Integrals and Partition Functions in Quantum Gravity". In: *Phys. Rev. D* 15 (1977), pp. 2752–2756. DOI: 10.1103/PhysRevD.15.2752.

[14] Juan Martin Maldacena. "The Large N limit of superconformal field theories and supergravity". In: *Adv. Theor. Math. Phys.* 2 (1998), pp. 231–252. DOI: 10.4310/ATMP.1998.v2.n2.a1. arXiv: hep-th/9711200.

[15] Edward Witten. "Anti-de Sitter space and holography". In: *Adv. Theor. Math. Phys.* 2 (1998), pp. 253–291. DOI: 10.4310/ATMP.1998.v2.n2.a2. arXiv: hep-th/9802150.

[16] S. S. Gubser, Igor R. Klebanov, and Alexander M. Polyakov. "Gauge theory correlators from noncritical string theory". In: *Phys. Lett. B* 428 (1998), pp. 105–114. DOI: 10.1016/S0370-2693(98)00377-3. arXiv: hep-th/9802109.

[17] Shinsei Ryu and Tadashi Takayanagi. "Holographic derivation of entanglement entropy from AdS/CFT". In: *Phys. Rev. Lett.* 96 (2006), p. 181602. DOI: 10.1103/PhysRevLett.96.181602. arXiv: hep-th/0603001.

[18] Veronika E. Hubeny, Mukund Rangamani, and Tadashi Takayanagi. "A Covariant holographic entanglement entropy proposal". In: *JHEP* 07 (2007), p. 062. DOI: 10.1088/1126-6708/2007/07/062. arXiv: 0705.0016 [hep-th].

[19] Thomas Faulkner, Aitor Lewkowycz, and Juan Maldacena. "Quantum corrections to holographic entanglement entropy". In: *JHEP* 11 (2013), p. 074. DOI: 10.1007/JHEP11(2013)074. arXiv: 1307.2892 [hep-th].

[20] Netta Engelhardt and Aron C. Wall. "Quantum Extremal Surfaces: Holographic Entanglement Entropy beyond the Classical Regime". In: *JHEP* 01 (2015), p. 073. DOI: 10.1007/JHEP01(2015)073. arXiv: 1408.3203 [hep-th].

[21] Seiji Terashima. *Simple Bulk Reconstruction in AdS/CFT Correspondence*. 2022. arXiv: 2104.11743 [hep-th]. URL: https://arxiv.org/abs/2104.11743.

[22] Seiji Terashima. "Bulk locality in the ¡mml:math xmlns:mml="http://www.w3.org/1998/Math/M display="inline"¿¡mml:mrow¿¡mml:mi¿AdS¡/mml:mi¿¡mml:mo¿/¡/mml:mo¿¡mml:mi¿CFT¡/mm correspondence". In: *Physical Review D* 104.8 (Oct. 2021). ISSN: 2470-0029. DOI: 10.1103/physrevd.104.086014. URL: http://dx.doi.org/10.1103/PhysRevD.104.086014.

[23] Sotaro Sugishita and Seiji Terashima. *Bulk Reconstruction and Gauge Invariance*. 2024. arXiv: 2409.02534 [hep-th]. URL: https://arxiv.org/abs/2409.02534.

[24] Aron C. Wall. "Maximin Surfaces, and the Strong Subadditivity of the Covariant Holographic Entanglement Entropy". In: *Class. Quant. Grav.* 31.22 (2014), p. 225007. DOI: 10.1088/0264-9381/31/22/225007. arXiv: 1211.3494 [hep-th].

[25] Matthew Headrick, Veronika E. Hubeny, Albion Lawrence, and Mukund Rangamani. "Causality & holographic entanglement entropy". In: *JHEP* 12 (2014), p. 162. DOI: 10.1007/JHEP12(2014)162. arXiv: 1408.6300 [hep-th].

[26] Bartlomiej Czech, Joanna L. Karczmarek, Fernando Nogueira, and Mark Van Raamsdonk. "The Gravity Dual of a Density Matrix". In: *Class. Quant. Grav.* 29 (2012), p. 155009. DOI: 10.1088/0264-9381/29/15/155009. arXiv: 1204.1330 [hep-th].

[27] Daniel L. Jafferis, Aitor Lewkowycz, Juan Maldacena, and S. Josephine Suh. "Relative entropy equals bulk relative entropy". In: *JHEP* 06 (2016), p. 004. DOI: 10.1007/JHEP06(2016)004. arXiv: 1512.06431 [hep-th].

[28] Xi Dong, Daniel Harlow, and Aron C. Wall. "Reconstruction of Bulk Operators within the Entanglement Wedge in Gauge-Gravity Duality". In: *Phys. Rev. Lett.* 117.2 (2016), p. 021601. DOI: 10.1103/PhysRevLett.117.021601. arXiv: 1601.05416 [hep-th].

[29] Jordan Cotler, Patrick Hayden, Geoffrey Penington, Grant Salton, Brian Swingle, and Michael Walter. "Entanglement Wedge Reconstruction via Universal Recovery Channels". In: *Phys. Rev. X* 9.3 (2019), p. 031011. DOI: 10.1103/PhysRevX.9.031011. arXiv: 1704.05839 [hep-th].

[30] Ahmed Almheiri, Xi Dong, and Daniel Harlow. "Bulk Locality and Quantum Error Correction in AdS/CFT". In: *JHEP* 04 (2015), p. 163. DOI: 10.1007/JHEP04(2015)163. arXiv: 1411.7041 [hep-th].

[31] Daniel Harlow. "The Ryu–Takayanagi Formula from Quantum Error Correction". In: *Commun. Math. Phys.* 354.3 (2017), pp. 865–912. DOI: 10.1007/s00220-017-2904-z. arXiv: 1607.03901 [hep-th].

[32] Xi Dong, Xiao-Liang Qi, and Michael Walter. "Holographic entanglement negativity and replica symmetry breaking". In: *JHEP* 06 (2021), p. 024. DOI: 10.1007/JHEP06(2021)024. arXiv: 2101.11029 [hep-th].

[33] Souvik Dutta and Thomas Faulkner. "A canonical purification for the entanglement wedge cross-section". In: *JHEP* 03 (2021), p. 178. DOI: 10.1007/JHEP03(2021)178. arXiv: 1905.00577 [hep-th].

[34] Phil Saad, Stephen H. Shenker, and Douglas Stanford. "A semiclassical ramp in SYK and in gravity". In: (June 2018). arXiv: 1806.06840 [hep-th].

[35] Phil Saad, Stephen H. Shenker, and Douglas Stanford. "JT gravity as a matrix integral". In: (Mar. 2019). arXiv: 1903.11115 [hep-th].

[36] Andreas Blommaert, Thomas G. Mertens, and Henri Verschelde. "Clocks and Rods in Jackiw-Teitelboim Quantum Gravity". In: *JHEP* 09 (2019), p. 060. DOI: 10.1007/JHEP09(2019)060. arXiv: 1902.11194 [hep-th].

[37] Phil Saad. "Late Time Correlation Functions, Baby Universes, and ETH in JT Gravity". In: (Oct. 2019). arXiv: 1910.10311 [hep-th].

[38] Andreas Blommaert. "Dissecting the ensemble in JT gravity". In: *JHEP* 09 (2022), p. 075. DOI: 10.1007/JHEP09(2022)075. arXiv: 2006.13971 [hep-th].

[39] Douglas Stanford, Zhenbin Yang, and Shunyu Yao. "Subleading Weingartens". In: *JHEP* 02 (2022), p. 200. DOI: 10.1007/JHEP02(2022)200. arXiv: 2107.10252 [hep-th].

[40] Juan Martin Maldacena. "Eternal black holes in anti-de Sitter". In: *JHEP* 04 (2003), p. 021. DOI: 10.1088/1126-6708/2003/04/021. arXiv: hep-th/0106112.

[41] Mark Srednicki. "Chaos and Quantum Thermalization". In: *Phys. Rev. E* 50 (Mar. 1994). DOI: 10.1103/PhysRevE.50.888. arXiv: cond-mat/9403051.

[42] J. M. Deutsch. "Quantum statistical mechanics in a closed system". In: *Phys. Rev. A* 43.4 (1991), p. 2046. DOI: 10.1103/PhysRevA.43.2046.

[43] R. Jackiw. "Lower Dimensional Gravity". In: *Nucl. Phys. B* 252 (1985). Ed. by R. Baier and H. Satz, pp. 343–356. DOI: 10.1016/0550-3213(85)90448-1.

[44] C. Teitelboim. "Gravitation and Hamiltonian Structure in Two Space-Time Dimensions". In: *Phys. Lett. B* 126 (1983), pp. 41–45. DOI: 10.1016/0370-2693(83)90012-6.

[45] A. Kitaev. *A simple model of quantum holography*. http://online.kitp.ucsb.edu/online/entangled15/kitaev/ and http://online.kitp.ucsb.edu/online/entangled15/kitaev2/. Talks at KITP, April 7, 2015 and May 27, 2015.

[46] Subir Sachdev and Jinwu Ye. "Gapless spin fluid ground state in a random, quantum Heisenberg magnet". In: *Phys. Rev. Lett.* 70 (1993), p. 3339. DOI: 10.1103/PhysRevLett.70.3339. arXiv: cond-mat/9212030.

[47] Patrick Hayden and John Preskill. "Black holes as mirrors: Quantum information in random subsystems". In: *JHEP* 09 (2007), p. 120. DOI: 10.1088/1126-6708/2007/09/120. arXiv: 0708.4025 [hep-th].

[48] Thomas G. Mertens and Gustavo J. Turiaci. "Solvable models of quantum black holes: a review on Jackiw–Teitelboim gravity". In: *Living Rev. Rel.* 26.1 (2023), p. 4. DOI: 10.1007/s41114-023-00046-1. arXiv: 2210.10846 [hep-th].

[49] Pranjal Nayak, Ashish Shukla, Ronak M. Soni, Sandip P. Trivedi, and V. Vishal. "On the Dynamics of Near-Extremal Black Holes". In: *JHEP* 09 (2018), p. 048. DOI: 10.1007/JHEP09(2018)048. arXiv: 1802.09547 [hep-th].

[50] Ahmed Almheiri and Joseph Polchinski. "Models of AdS_2 backreaction and holography". In: *JHEP* 11 (2015), p. 014. DOI: 10.1007/JHEP11(2015)014. arXiv: 1402.6334 [hep-th].

[51] S. M. Christensen and S. A. Fulling. "Trace anomalies and the Hawking effect". In: *Phys. Rev. D* 15 (8 1977), pp. 2088–2104. DOI: 10.1103/PhysRevD.15.2088. URL: https://link.aps.org/doi/10.1103/PhysRevD.15.2088.

[52] J. B. Hartle and S. W. Hawking. "Wave Function of the Universe". In: *Phys. Rev. D* 28 (1983). Ed. by Li-Zhi Fang and R. Ruffini, pp. 2960–2975. DOI: 10.1103/PhysRevD.28.2960.

[53] Ryogo Kubo. "Statistical mechanical theory of irreversible processes. 1. General theory and simple applications in magnetic and conduction problems". In: *J. Phys. Soc. Jap.* 12 (1957), pp. 570–586. DOI: 10.1143/JPSJ.12.570.

[54] Paul C. Martin and Julian Schwinger. "Theory of Many-Particle Systems. I". In: *Phys. Rev.* 115 (6 1959), pp. 1342–1373. DOI: 10.1103/PhysRev.115.1342. URL: https://link.aps.org/doi/10.1103/PhysRev.115.1342.

[55] Timothy J. Hollowood and S. Prem Kumar. "Islands and Page Curves for Evaporating Black Holes in JT Gravity". In: *JHEP* 08 (2020), p. 094. DOI: 10.1007/JHEP08(2020)094. arXiv: 2004.14944 [hep-th].

[56] Ahmed Almheiri, Raghu Mahajan, and Juan Maldacena. "Islands outside the horizon". In: (Oct. 2019). arXiv: 1910.11077 [hep-th].

[57] Timothy J. Hollowood, S. Prem Kumar, Andrea Legramandi, and Neil Talwar. "Islands in the stream of Hawking radiation". In: *JHEP* 11 (2021), p. 067. DOI: 10.1007/JHEP11(2021)067. arXiv: 2104.00052 [hep-th].

[58] Timothy J. Hollowood, S. Prem Kumar, Andrea Legramandi, and Neil Talwar. "Grey-body factors, irreversibility and multiple island saddles". In: *JHEP* 03 (2022), p. 110. DOI: 10.1007/JHEP03(2022)110. arXiv: 2111.02248 [hep-th].

[59] Zsolt Gyongyosi, Timothy J. Hollowood, S. Prem Kumar, Andrea Legramandi, and Neil Talwar. "Black Hole Information Recovery in JT Gravity". In: *JHEP* 01 (2023), p. 139. DOI: 10.1007/JHEP01(2023)139. arXiv: 2209.11774 [hep-th].

[60] Ahmed Almheiri, Raghu Mahajan, Juan Maldacena, and Ying Zhao. "The Page curve of Hawking radiation from semiclassical geometry". In: *Journal of High Energy Physics* 2020.3 (Mar. 2020). ISSN: 1029-8479. DOI: 10.1007/jhep03(2020)149. URL: http://dx.doi.org/10.1007/JHEP03(2020)149.

[61] Raphael Bousso and Arvin Shahbazi-Moghaddam. "Island finder and entropy bound". In: *Physical Review D* 103.10 (May 2021). ISSN: 2470-0029. DOI: 10.1103/physrevd.103.106005. URL: http://dx.doi.org/10.1103/PhysRevD.103.106005.

[62] Raphael Bousso and Elizabeth Wildenhain. "Gravity/ensemble duality". In: *Physical Review D* 102.6 (Sept. 2020). ISSN: 2470-0029. DOI: 10.1103/physrevd.102.066005. URL: http://dx.doi.org/10.1103/PhysRevD.102.066005.

[63] Kanato Goto, Thomas Hartman, and Amirhossein Tajdini. "Replica wormholes for an evaporating 2D black hole". In: *Journal of High Energy Physics* 2021.4 (Apr. 2021). ISSN: 1029-8479. DOI: 10.1007/jhep04(2021)289. URL: http://dx.doi.org/10.1007/JHEP04(2021)289.

[64] Yoshinori Matsuo. "Islands and stretched horizon". In: *Journal of High Energy Physics* 2021.7 (July 2021). ISSN: 1029-8479. DOI: 10.1007/jhep07(2021)051. URL: http://dx.doi.org/10.1007/JHEP07(2021)051.

[65] Hong Zhe Chen, Robert C. Myers, Dominik Neuenfeld, Ignacio A. Reyes, and Joshua Sandor. *Quantum Extremal Islands Made Easy, Part I: Entanglement on the Brane*. 2020. arXiv: 2006.04851 [hep-th]. URL: https://arxiv.org/abs/2006.04851.

[66] Yi Ling, Yuxuan Liu, and Zhuo-Yu Xian. "Island in charged black holes". In: *Journal of High Energy Physics* 2021.3 (Mar. 2021). ISSN: 1029-8479. DOI: 10.1007/jhep03(2021)251. URL: http://dx.doi.org/10.1007/JHEP03(2021)251.

[67] Timothy J. Hollowood and S. Prem Kumar. *Islands and Page Curves for Evaporating Black Holes in JT Gravity*. 2020. arXiv: 2004.14944 [hep-th]. URL: https://arxiv.org/abs/2004.14944.

[68] Timothy J Hollowood, S Prem Kumar, and Andrea Legramandi. "Hawking radiation correlations of evaporating black holes in JT gravity". In: *Journal of Physics A: Mathematical and Theoretical* 53.47 (Oct. 2020), p. 475401. ISSN: 1751-8121. DOI: 10.1088/1751-8121/abbc51. URL: http://dx.doi.org/10.1088/1751-8121/abbc51.

[69] Hong Zhe Chen, Zachary Fisher, Juan Hernandez, Robert C. Myers, and Shan-Ming Ruan. *Evaporating Black Holes Coupled to a Thermal Bath*. 2020. arXiv: 2007.11658 [hep-th]. URL: https://arxiv.org/abs/2007.11658.

[70] Yiming Chen. "Pulling out the island with modular flow". In: *Journal of High Energy Physics* 2020.3 (Mar. 2020). ISSN: 1029-8479. DOI: 10.1007/jhep03(2020)033. URL: http://dx.doi.org/10.1007/JHEP03(2020)033.

[71] Chethan Krishnan, Vaishnavi Patil, and Jude Pereira. *Page Curve and the Information Paradox in Flat Space*. 2020. arXiv: 2005.02993 [hep-th]. URL: https://arxiv.org/abs/2005.02993.

[72] Alok Laddha, Siddharth G. Prabhu, Suvrat Raju, and Pushkal Shrivastava. "The Holographic Nature of Null Infinity". In: *SciPost Phys.* 10.2 (2021), p. 041. DOI: 10.21468/SciPostPhys.10.2.041. arXiv: 2002.02448 [hep-th].

[73] Hao Geng, Andreas Karch, Carlos Perez-Pardavila, Suvrat Raju, Lisa Randall, Marcos Riojas, and Sanjit Shashi. "Information Transfer with a Gravitating Bath". In: *SciPost Phys.* 10.5 (2021), p. 103. DOI: 10.21468/SciPostPhys.10.5.103. arXiv: 2012.04671 [hep-th].

[74] Hao Geng, Andreas Karch, Carlos Perez-Pardavila, Suvrat Raju, Lisa Randall, Marcos Riojas, and Sanjit Shashi. "Inconsistency of islands in theories with long-range gravity". In: *JHEP* 01 (2022), p. 182. DOI: 10.1007/JHEP01(2022)182. arXiv: 2107.03390 [hep-th].

[75] Suvrat Raju. "Failure of the split property in gravity and the information paradox". In: *Class. Quant. Grav.* 39.6 (2022), p. 064002. DOI: 10.1088/1361-6382/ac482b. arXiv: 2110.05470 [hep-th].

[76] Hao Geng, Andreas Karch, Carlos Perez-Pardavila, Suvrat Raju, Lisa Randall, Marcos Riojas, and Sanjit Shashi. "Entanglement phase structure of a holographic BCFT in a black hole background". In: *JHEP* 05 (2022), p. 153. DOI: 10.1007/JHEP05(2022)153. arXiv: 2112.09132 [hep-th].

[77] Emil J. Martinec. "AdS3's with and without BTZ's". In: (Sept. 2021). arXiv: 2109.11716 [hep-th].

[78] Emil J. Martinec. "Trouble in Paradox". In: (Mar. 2022). arXiv: 2203.04947 [hep-th].

[79] H. Casini, C. D. Fosco, and M. Huerta. "Entanglement and alpha entropies for a massive Dirac field in two dimensions". In: *J. Stat. Mech.* 0507 (2005), P07007. DOI: 10.1088/1742-5468/2005/07/P07007. arXiv: cond-mat/0505563.

[80] N. D. Birrell and P. C. W. Davies. *Quantum Fields in Curved Space*. Cambridge Monographs on Mathematical Physics. Cambridge, UK: Cambridge Univ. Press, Feb. 1984. ISBN: 978-0-521-27858-4, 978-0-521-27858-4. DOI: 10.1017/CBO9780511622632.

[81] Jacob D. Bekenstein. "Generalized second law of thermodynamics in black-hole physics". In: *Phys. Rev. D* 9 (12 1974), pp. 3292–3300. DOI: 10.1103/PhysRevD.9.3292. URL: https://link.aps.org/doi/10.1103/PhysRevD.9.3292.

[82] Patrick Hayden and Geoffrey Penington. "Learning the Alpha-bits of Black Holes". In: *JHEP* 12 (2019), p. 007. DOI: 10.1007/JHEP12(2019)007. arXiv: 1807.06041 [hep-th].

[83] H. Araki and E. H. Lieb. "Entropy inequalities". In: *Commun. Math. Phys.* 18 (1970), pp. 160–170. DOI: 10.1007/BF01646092.

[84] Chris Akers, Netta Engelhardt, Geoff Penington, and Mykhaylo Usatyuk. "Quantum Maximin Surfaces". In: *JHEP* 08 (2020), p. 140. DOI: 10.1007/JHEP08(2020)140. arXiv: 1912.02799 [hep-th].

[85] Matthias Christandl and Andreas Winter. ""Squashed entanglement": An additive entanglement measure". In: *J. Math. Phys.* 45.3 (2004), p. 829. DOI: 10.1063/1.1643788.

[86] Anura Abeyesinghe, Igor Devetak, Patrick Hayden, and Andreas Winter. "The mother of all protocols: Restructuring quantum information's family tree". In: *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 465.2108 (2009), pp. 2537–2563.

[87] Donald Marolf and Henry Maxfield. "Observations of Hawking radiation: the Page curve and baby universes". In: *JHEP* 04 (2021), p. 272. DOI: 10.1007/JHEP04(2021)272. arXiv: 2010.06602 [hep-th].

[88] Jonah Kudler-Flam, Vladimir Narovlansky, and Shinsei Ryu. "Negativity spectra in random tensor networks and holography". In: *JHEP* 02 (2022), p. 076. DOI: 10.1007/JHEP02(2022)076. arXiv: 2109.02649 [hep-th].

[89] Daniel Harlow and Patrick Hayden. "Quantum Computation vs. Firewalls". In: *JHEP* 06 (2013), p. 085. DOI: 10.1007/JHEP06(2013)085. arXiv: 1301.4504 [hep-th].

[90] Jacob D. Bekenstein. "A Universal Upper Bound on the Entropy to Energy Ratio for Bounded Systems". In: *Phys. Rev. D* 23 (1981), p. 287. DOI: 10.1103/PhysRevD.23.287.

[91] H. Casini. "Relative entropy and the Bekenstein bound". In: *Class. Quant. Grav.* 25 (2008), p. 205021. DOI: 10.1088/0264-9381/25/20/205021. arXiv: 0804.2182 [hep-th].

[92] Denes Petz. "SUFFICIENCY OF CHANNELS OVER VON NEUMANN ALGEBRAS". In: *Quart. J. Math. Oxford Ser.* 39.1 (1988), pp. 97–108. DOI: 10.1093/qmath/39.1.97.

[93] Adam R. Brown, Hrant Gharibyan, Geoff Penington, and Leonard Susskind. "The Python's Lunch: geometric obstructions to decoding Hawking radiation". In: *JHEP* 08 (2020), p. 121. DOI: 10.1007/JHEP08(2020)121. arXiv: 1912.00228 [hep-th].

[94] Patrick Hayden and Geoffrey Penington. "Approximate Quantum Error Correction Revisited: Introducing the Alpha-Bit". In: *Commun. Math. Phys.* 374.2 (2020), pp. 369–432. DOI: 10.1007/s00220-020-03689-1. arXiv: 1706.09434 [quant-ph].

[95] Pasquale Calabrese and John L. Cardy. "Evolution of entanglement entropy in one-dimensional systems". In: *J. Stat. Mech.* 0504 (2005), P04010. DOI: 10.1088/1742-5468/2005/04/P04010. arXiv: cond-mat/0503393.

[96] Masahiro Nozaki, Tokiro Numasawa, and Tadashi Takayanagi. "Quantum Entanglement of Local Operators in Conformal Field Theories". In: *Phys. Rev. Lett.* 112 (2014), p. 111602. DOI: 10.1103/PhysRevLett.112.111602. arXiv: 1401.0539 [hep-th].

[97] Pawel Caputa, Masahiro Nozaki, and Tadashi Takayanagi. "Entanglement of local operators in large-N conformal field theories". In: *PTEP* 2014 (2014), 093B06. DOI: 10.1093/ptep/ptu122. arXiv: 1405.5946 [hep-th].

[98] Beni Yoshida and Alexei Kitaev. "Efficient decoding for the Hayden-Preskill protocol". In: (Oct. 2017). arXiv: 1710.03363 [hep-th].

[99] Chris Akers, Netta Engelhardt, Daniel Harlow, Geoff Penington, and Shreya Vardhan. "The black hole interior from non-isometric codes and complexity". In: *JHEP* 06 (2024), p. 155. DOI: 10.1007/JHEP06(2024)155. arXiv: 2207.06536 [hep-th].

[100] Isaac H. Kim and John Preskill. "Complementarity and the unitarity of the black hole S-matrix". In: *JHEP* 02 (2023), p. 233. DOI: 10.1007/JHEP02(2023)233. arXiv: 2212.00194 [hep-th].

[101] Chi-Fang Chen, Geoffrey Penington, and Grant Salton. "Entanglement Wedge Reconstruction using the Petz Map". In: *JHEP* 01 (2020), p. 168. DOI: 10.1007/JHEP01(2020)168. arXiv: 1902.02844 [hep-th].

[102] Chris Akers and Pratik Rath. "Holographic Renyi Entropy from Quantum Error Correction". In: *JHEP* 05 (2019), p. 052. DOI: 10.1007/JHEP05(2019)052. arXiv: 1811.05171 [hep-th].

[103] Xi Dong, Daniel Harlow, and Donald Marolf. "Flat entanglement spectra in fixed-area states of quantum gravity". In: *JHEP* 10 (2019), p. 240. DOI: 10.1007/JHEP10(2019)240. arXiv: 1811.05382 [hep-th].

[104] Chris Akers and Geoff Penington. "Leading order corrections to the quantum extremal surface prescription". In: *JHEP* 04 (2021), p. 062. DOI: 10.1007/JHEP04(2021)062. arXiv: 2008.03319 [hep-th].

[105] Jinzhao Wang. "The refined quantum extremal surface prescription from the asymptotic equipartition property". In: *Quantum* 6 (2022), p. 655. DOI: 10.22331/q-2022-02-16-655. arXiv: 2105.05892 [hep-th].

[106] Timothy J. Hollowood, S. Prem Kumar, Andrea Legramandi, and Neil Talwar. "Ephemeral islands, plunging quantum extremal surfaces and BCFT channels". In: *JHEP* 01 (2022), p. 078. DOI: 10.1007/JHEP01(2022)078. arXiv: 2109.01895 [hep-th].

[107] Samir D. Mathur. "What is the dual of two entangled CFTs?" In: (Feb. 2014). arXiv: 1402.6378 [hep-th].

[108] Curtis T. Asplund, Alice Bernamonti, Federico Galli, and Thomas Hartman. "Entanglement Scrambling in 2d Conformal Field Theory". In: *JHEP* 09 (2015), p. 110. DOI: 10.1007/JHEP09(2015)110. arXiv: 1506.03772 [hep-th].

[109] Tadashi Takayanagi. "Holographic Dual of BCFT". In: *Phys. Rev. Lett.* 107 (2011), p. 101602. DOI: 10.1103/PhysRevLett.107.101602. arXiv: 1105.5165 [hep-th].

[110] Vijay Balasubramanian, Ben Craps, Mikhail Khramtsov, and Edgar Shaghoulian. "Submerging islands through thermalization". In: *JHEP* 10 (2021), p. 048. DOI: 10.1007/JHEP10(2021)048. arXiv: 2107.14746 [hep-th].

[111] John L. Cardy. "Conformal Invariance and Surface Critical Behavior". In: *Nucl. Phys. B* 240 (1984), pp. 514–532. DOI: 10.1016/0550-3213(84)90241-4.

[112] P. Di Francesco, P. Mathieu, and D. Senechal. *Conformal Field Theory*. Graduate Texts in Contemporary Physics. New York: Springer-Verlag, 1997. ISBN: 978-0-387-94785-3, 978-1-4612-7475-9. DOI: 10.1007/978-1-4612-2256-9.

[113] John L. Cardy. "Boundary conformal field theory". In: (Nov. 2004). arXiv: hep-th/0411189.

[114] Ingo Runkel. "Boundary Problems in Conformal Field Theory". PhD thesis. King's College London, University of London, 2000. URL: https://www.math.uni-hamburg.de/home/runkel/PDF/phd.pdf.

[115] Ian Affleck and Andreas W. W. Ludwig. "Universal noninteger 'ground state degeneracy' in critical quantum systems". In: *Phys. Rev. Lett.* 67 (1991), pp. 161–164. DOI: 10.1103/PhysRevLett.67.161.

[116] Ahmed Almheiri, Raghu Mahajan, and Jorge E. Santos. "Entanglement islands in higher dimensions". In: *SciPost Phys.* 9.1 (2020), p. 001. DOI: 10.21468/SciPostPhys.9.1.001. arXiv: 1911.09666 [hep-th].

[117] Hao Geng, Severin Lüst, Rashmish K. Mishra, and David Wakeham. "Holographic BCFTs and Communicating Black Holes". In: *jhep* 08 (2021), p. 003. DOI: 10.1007/JHEP08(2021)003. arXiv: 2104.07039 [hep-th].

[118] Thomas Hartman. "Entanglement Entropy at Large Central Charge". In: (Mar. 2013). arXiv: 1303.6955 [hep-th].

[119] James Sully, Mark Van Raamsdonk, and David Wakeham. "BCFT entanglement entropy at large central charge and the black hole interior". In: *JHEP* 03 (2021), p. 167. DOI: 10.1007/JHEP03(2021)167. arXiv: 2004.13088 [hep-th].

[120] J. David Brown and M. Henneaux. "Central Charges in the Canonical Realization of Asymptotic Symmetries: An Example from Three-Dimensional Gravity". In: *Commun. Math. Phys.* 104 (1986), pp. 207–226. DOI: 10.1007/BF01211590.

[121] Moshe Rozali, James Sully, Mark Van Raamsdonk, Christopher Waddell, and David Wakeham. "Information radiation in BCFT models of black holes". In: *JHEP* 05 (2020), p. 004. DOI: 10.1007/JHEP05(2020)004. arXiv: 1910.12836 [hep-th].

[122] O. Bohigas, M. J. Giannoni, and C. Schmit. "Characterization of chaotic quantum spectra and universality of level fluctuation laws". In: *Phys. Rev. Lett.* 52 (1984), pp. 1–4. DOI: 10.1103/PhysRevLett.52.1.

[123] Juan Maldacena and Douglas Stanford. "Remarks on the Sachdev-Ye-Kitaev model". In: *Phys. Rev. D* 94.10 (2016), p. 106002. DOI: 10.1103/PhysRevD.94.106002. arXiv: 1604.07818 [hep-th].

[124] Michael Winer, Shao-Kai Jian, and Brian Swingle. "An exponential ramp in the quadratic Sachdev-Ye-Kitaev model". In: *Phys. Rev. Lett.* 125 (2020), p. 250602. DOI: 10.1103/PhysRevLett.125.250602. arXiv: 2006.15152 [cond-mat.stat-mech].

[125] Phil Saad, Stephen H. Shenker, Douglas Stanford, and Shunyu Yao. "Wormholes without averaging". In: (Mar. 2021). arXiv: 2103.16754 [hep-th].

[126] Dionysios Anninos, Tarek Anous, Ben Pethybridge, and Gizem Şengör. "The discreet charm of the discrete series in dS_2". In: *J. Phys. A* 57.2 (2024), p. 025401. DOI: 10.1088/1751-8121/ad14ad. arXiv: 2307.15832 [hep-th].

[127] E. Brezin and S. Hikami. "Intersection theory from duality and replica". In: *Commun. Math. Phys.* 283 (2008), pp. 507–521. DOI: 10.1007/s00220-008-0519-0. arXiv: 0708.2210 [hep-th].

[128] Shenglong Xu, Leonard Susskind, Yuan Su, and Brian Swingle. "A Sparse Model of Quantum Holography". In: (Aug. 2020). arXiv: 2008.02303 [cond-mat.str-el].

[129] Antonio M. García-García, Bruno Loureiro, Aurelio Romero-Bermúdez, and Masaki Tezuka. "Chaotic-Integrable Transition in the Sachdev-Ye-Kitaev Model". In: *Phys. Rev. Lett.* 120.24 (2018), p. 241603. DOI: 10.1103/PhysRevLett.120.241603. arXiv: 1707.02197 [hep-th].

[130] Chris Akers, Thomas Faulkner, Simon Lin, and Pratik Rath. "Reflected entropy in random tensor networks". In: *JHEP* 05 (2022), p. 162. DOI: 10.1007/JHEP05(2022)162. arXiv: 2112.09122 [hep-th].