Replica wormholes and the black hole interior

Geoff Penington, Stephen H. Shenker, Douglas Stanford and Zhenbin Yang

Stanford Institute for Theoretical Physics, Stanford University,
Stanford, CA 94305, U.S.A.

E-mail: geoffp@berkeley.edu, sshenker@stanford.edu,
salguod@stanford.edu, zhenbiny@stanford.edu

ABSTRACT: Recent work has shown how to obtain the Page curve of an evaporating black hole from holographic computations of entanglement entropy. We show how these computations can be justified using the replica trick, from geometries with a spacetime wormhole connecting the different replicas. In a simple model, we study the Page transition in detail by summing replica geometries with different topologies. We compute related quantities in less detail in more complicated models, including JT gravity coupled to conformal matter and the SYK model. Separately, we give a direct gravitational argument for entanglement wedge reconstruction using an explicit formula known as the Petz map; again, a spacetime wormhole plays an important role. We discuss an interpretation of the wormhole geometries as part of some ensemble average implicit in the gravity description.

Contents

1 Introduction

Hawking's calculation [1] of black hole evaporation by thermal radiation raises deep puzzles about the consistency of black holes and quantum mechanics. In particular, if the black hole starts in a pure quantum state, how can the apparently thermal radiation at the end of evaporation actually be in a pure state, as quantum mechanics requires? A quantitative form of this puzzle is encapsulated in the Page curve [2], a plot of the entanglement entropy of the radiation as a function of time. It increases at short times because of the thermal character of the radiation, but then must decrease to zero at the end of evaporation to be consistent with unitary quantum mechanical evolution.

Holographic duality has supplied a powerful tool for computing entanglement entropy, the Ryu-Takayanagi formula [3]. This formula has been generalized and refined in numerous ways since its original formulation.[1] In its most general formulation, the Engelhardt-Wall (EW) prescription [11], it says that the entropy of a holographic boundary region B is given by the generalized entropy of the minimal quantum extremal surface (QES).

Building on earlier ideas in [12, 13], the authors of [14, 15] considered a black hole in anti-de Sitter space that was allowed to evaporate into an auxiliary system R using absorbing boundary conditions. They showed that, at exactly the Page time, there was a phase transition in the minimal QES. This caused the boundary entropy to begin decreasing, in accordance with the Page curve. However, as emphasized in [15], a naive calculation of the entropy of the Hawking radiation showed that it continued to increase, since the semiclassical bulk physics had not changed.

As advocated in [14], and commented on in [15], the Page curve for the radiation would result from a variant of the usual rules for computing entropy holographically.[2] This idea was given an elegant "doubly holographic" realization in [16], where it was called the "island conjecture." Explicitly, the prescription states that the actual entropy $S(R)$ of the Hawking radiation is given by

$$S(R) = \min\left\{ \underset{I}{\text{ext}}\left[\frac{\text{Area}(\partial I)}{4G_N} + S_{\text{bulk}}(I \cup R) \right] \right\}, \tag{1.1}$$

where I is some "island" region of the bulk.[3] Here $S_{\text{bulk}}(I \cup R)$ is the entropy of the island plus the Hawking radiation, computed semiclassically in the original fixed geometry.[4]

---

[1]For example [4–9]; for a review see [10].

[2]This is a special case of a prescription introduced in [13] (near eq. 4.14) for the entropy of the *combination* of a boundary region B and auxiliary system R. If B is empty, the prescription reduces to (1.1).

[3]In [13, 14], (1.1) was justified by imagining throwing the auxiliary, nonholographic system R into an auxiliary *holographic* system. (1.1) then reduces to the usual EW prescription. In [16], (1.1) was justified by considering holographic bulk matter and using the Ryu-Takayanagi formula in the "doubly holographic" description of the theory.

[4]For other recent work on this topic, see [17–22].

Since the Ryu-Takayanagi formula can be derived from gravitational path integrals using replicas [7–9, 23], the prescription (1.1) should also be derivable directly from the gravitational path integral. Deriving it is one of the main goals of this paper. We will show explicitly how this prescription arises from a replica computation with Euclidean wormholes connecting the different replicas. While we present our arguments in the context of simple models, their structure is not model dependent and should apply in a general context.

Wormholes of a similar type have recently been used to analyze the late-time behavior of the spectral form factor [24, 25] and correlation functions [26], inspired by the puzzle posed in [27]. In both that setting and here, the basic point of the wormholes is to give small but nonzero overlaps between naively orthogonal bulk states. This resonates with the longstanding suspicion that some small corrections to the Hawking calculation (of order $e^{-S_{\mathrm{BH}}}$) could be responsible for making black hole evolution unitary.[5]

Understanding the Page curve is only part of the black hole information problem. One also wants to understand how information that was thrown into the black hole ends up escaping in the Hawking radiation. As argued in [14, 15], this can be addressed using *entanglement wedge reconstruction*.[6] In the case of the island conjecture, this idea implies that R contains all of the information in the island I, which itself contains much of the black hole interior. Existing derivations of entanglement wedge reconstruction have been indirect,[7] relying on throwing entropy calculations into the meat grinder of modern quantum information theory. So, the second main goal of this paper is to show using a bulk argument that operators in the radiation can manipulate the interior of the black hole. We show this directly by using gravitational path integrals to evaluate matrix elements of an explicit reconstruction operator defined using the Petz map [41, 42] (see [39, 43] for a description and recent discussion of the Petz map). Again, Euclidean wormholes play a crucial role. Here they are in some sense connecting the interior of the black hole to the quantum computer acting on the Hawking radiation.

We now give a brief summary of this paper.

In section 2 we introduce a simple toy model of an evaporating black hole in Jackiw-Teitelboim (JT) gravity [44–46], a two dimensional truncation of near extremal black hole dynamics. We represent the Hawking radiation by an auxiliary reference system whose states are entangled with interior partner modes represented by "end of the world branes" (EOW branes) in the black hole interior. This model is simple enough that we can exactly evaluate the full gravitational path integral for the Rényi entropies by summing over all planar topologies with the correct boundary conditions. We find that, before the analog of the Page time, the dominant topology consists of n disconnected copies, one for each replica, of the single replica geometry. In contrast, after the Page time, the dominant topology is connected, with an n-boundary Euclidean wormhole connecting all the different replicas.

---

[5]A challenge to this idea was provided by [28, 29]: if the early radiation, the next Hawking quantum, and its interior partner are independent systems, the needed corrections would require a firewall at the horizon. However, if the interior partner is secretly part of the early radiation, then the arguments of [28, 29] do not apply [30–34]. The ER=EPR proposal [34] suggested that a geometric connection might play a role in this identification.

[6]This was originally conjectured in [6, 35, 36], and established in [37–39] using the ideas of [8].

[7]A partial exception is [40], but it does not immediately apply to the case of an evaporating black hole.

Our calculation uses a type of Feynman-diagram resummation method applied to spacetime geometry. This is inspired by techniques from the theory of free probability, and it makes it possible to continue the sum over topologies in n.

We also compute the overlap between the different black hole microstates $|\psi_i\rangle$. We find that $\langle\psi_i|\psi_j\rangle = 0$ for $i \neq j$, but that $|\langle\psi_i|\psi_j\rangle|^2 \sim e^{-S}$ due to a wormhole contribution. This can only be consistent if the gravitational path integral represents an ensemble of quantum theories.[8] The Rényi entropies are sums of large numbers of these overlaps. The Page transition comes from the buildup of these small errors in orthogonality.

In section 3 we explicitly reconstruct the black hole interior from the radiation in this simple model. Our tool is the Petz map which we introduce and explain. We then evaluate matrix elements of Petz map reconstructions using gravitational path integrals. We employ an analytic continuation similar to the replica trick and so consider geometries of different topologies. After analytic calculation, the problem reduces to a bulk field theory calculation in the original spacetime geometry.

After the Page time, when the connected topology dominates, the matrix elements of the Petz map reconstruction and original interior operator agree. In contrast, before the Page time, the disconnected topology dominates and the Petz map reconstruction fails to learn anything about the interior. Again we are able to sum these topologies and thereby follow this transition in detail. The error in the reconstruction increases linearly with the dimension of the code subspace. As a result, even long after the Page time, reconstruction is only possible for sufficiently small code subspaces, an example of state dependence. This explains the state dependence required in the Hayden-Preskill decoding process [47] and gives a direct derivation of the results about state dependence in entanglement wedge reconstruction from [13].

The toy model calculations have some simplifying features that actual calculations for evaporating black holes do not. We therefore extend our calculations to two less idealized models.

In section 4 we calculate an analog of the Page curve using the replica trick in JT gravity coupled to conformal matter, where the matter sector is allowed to flow freely across the boundaries between two thermofield double black holes. Perturbative semiclassical gravity would suggest that the entropy of each thermofield double black hole would grow forever. In contrast, unitarity and more explicitly the island formula (1.1) implies that it must saturate at twice the Bekenstein-Hawking entropy. We analyze this situation using replicas as before. In this case, we cannot explicitly find the saddle point for integer n Rényi entropies. But we are able to calculate the von Neumann entropy by showing that, as $n \to 1$, the dominant saddle point after the Page time replicates about the minimal QES. This reproduces the answer from (1.1). This argument essentially reproduces the general arguments for the EW prescription from [23] in this specific context.

In section 5, we do essentially the same calculations in a UV-complete theory, the SYK model. In this case, we can numerically find a replica-nondiagonal saddle point in the G, Σ

---

[8]For this particular model this is indeed the case. In appendix D, we construct an explicit ensemble of theories dual to our bulk model.
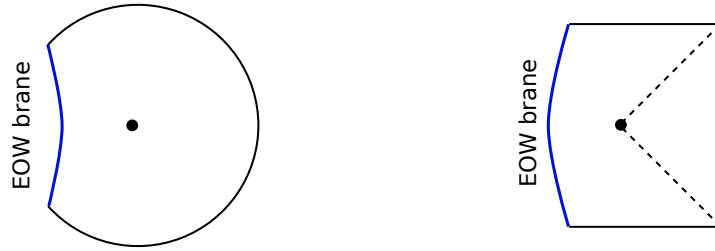
Figure 1. Euclidean and Lorentzian geometries for a black hole with an EOW brane behind the horizon.

action that describes the saturation of the $n = 2$ Rényi entropy. According to the usual intuition in this context for associating topologies to saddles for the G, Σ action [24], this saddle point corresponds to a Euclidean wormhole that connects the two replicas, as in our gravity calculations.

In section 6 we turn briefly to de Sitter space. We discuss a generalization of the no-boundary proposal for replica computations, and use it to find a Page-like transition in the entropy in a toy model using EOW branes in the JT gravity realization of de Sitter space [48]. A notable feature is that the density matrix itself changes character at the analog of the Page transition.

In section 7 we offer some preliminary remarks about the role of wormholes in theories without ensemble averaging.

We end with several appendices: in appendix A we discuss details of the n-replica gravitational partition function. In appendix B we present analogous calculations in more general theories of two dimensional dilaton gravity, including flat space. In appendix C we give a direct gravitational path integral derivation of general entanglement wedge reconstruction using the Petz map. In appendix D we present an explicit Hilbert space ensemble precisely dual to the sum over all topologies of the simple model. In appendix E we discuss in some detail the connection between random tensor networks and fixed area states. In appendix F we present a more refined analysis of the Page transition in the simple model.

Closely related work has been done independently by Ahmed Almheiri, Tom Hartman, Juan Maldacena, Edgar Shaghoulian and Amirhossein Tajdini [49]. We have arranged with these authors to coordinate submission of our papers.

2   A simple model

2.1   Setup of the model

In this section, we will introduce a simple 2d gravity model in which one can derive the "island" prescription (and corrections to it) in a very explicit way using the replica trick.[9] The starting point is to consider a black hole in JT gravity, with an "end of the world brane" (EOW brane) behind the horizon, sketched in figure 1. This can be understood as a $\mathbb{Z}_2$ quotient of an ordinary two-sided black hole, with a particle of mass μ behind the

---

[9]A similar model has recently been discussed independently in [19].

horizon [50]. The Euclidean action for the system is

$$I = I_{\mathsf{JT}} + \mu \int_{\text{brane}} \mathrm{d}s, \qquad (2.1)$$

where the integral is along the worldline of the EOW brane, and the pure JT action is

$$I_{\mathsf{JT}} = -\frac{S_0}{2\pi}\left[\frac{1}{2}\int_{\mathcal{M}}\sqrt{g}R + \int_{\partial\mathcal{M}}\sqrt{h}K\right] - \left[\frac{1}{2}\int_{\mathcal{M}}\sqrt{g}\phi(R+2) + \int_{\partial\mathcal{M}}\sqrt{h}\phi K\right]. \qquad (2.2)$$

In our discussion of the model, two different types of boundary conditions will be relevant. At the standard asymptotic boundary, we impose

$$\mathrm{d}s^2|_{\partial\mathcal{M}} = \frac{1}{\epsilon^2}\mathrm{d}\tau^2, \qquad \phi = \frac{1}{\epsilon}, \qquad \epsilon \to 0. \qquad (2.3)$$

Here τ can be interpreted as the imaginary time coordinate of the holographic boundary dual to the JT gravity system. At the EOW brane, we impose the "dual" boundary conditions

$$\partial_n\phi = \mu, \qquad K = 0, \qquad (2.4)$$

where ∂_n means the derivative normal to the EOW brane boundary, and $\mu \geq 0$.

We will be interested in the case where the EOW brane has a very large number k of possible internal states, each orthogonal to the others. To model an evaporating black hole, we will think of these states as describing the interior partners of the early Hawking radiation. We will entangle them with an auxiliary system R, which will model the early radiation of an evaporating black hole. So, all together, the state of the whole system is

$$|\Psi\rangle = \frac{1}{\sqrt{k}}\sum_{i=1}^{k}|\psi_i\rangle_{\mathsf{B}}|i\rangle_{\mathsf{R}}. \qquad (2.5)$$

Here $|\psi_i\rangle_{\mathsf{B}}$ is the state of the black hole B with the brane in state i, and $|i\rangle_{\mathsf{R}}$ is a state of the auxiliary "radiation" system R.

Let's compute the entropy of the R system in the state (2.5), using the island conjecture (1.1) of [14–16]. If we take the island to be the empty set, then the answer for the entropy is just $\log(k)$, representing the entanglement between the R system and the bulk state of the brane. However, we can also consider the case with a nontrivial island, as shown in figure 2. This island contains the EOW brane, so the bulk entropy term $S_{\text{bulk}}(\mathsf{R}\cup\mathsf{I})$ in the island formula (1.1) will give zero. The remaining area term is interpreted in JT gravity as

$$\frac{\text{Area}(\partial I)}{4G_N} \to S_0 + 2\pi\,\phi(\partial I). \qquad (2.6)$$

So the extremization in (1.1) amounts to putting the boundary of the island at an extremal point of the dilaton ϕ. The geometry shown in figure 1 has an extremal point of the dilaton, represented by the black dot at the bifurcation surface. The extremal value of $S_0 + 2\pi\phi$ is simply the coarse-grained entropy the black hole, S_{BH}. So, all together, the island conjecture predicts

$$S(\mathsf{R}) = \min\{\log(k), S_{\text{BH}}\}. \qquad (2.7)$$

The transition between these two answers, as a function of k, is analogous to the Page curve.

Figure 2. We show the expected entanglement wedges for our simple model based on the "island" conjecture. In the $k \ll e^{S_{\mathrm{BH}}}$ phase, the entanglement wedge of the B system (boundary dual of the gravity theory) is the whole spacetime, shown hatched. In the $k \gg e^{S_{\mathrm{BH}}}$ phase, an island develops. The entanglement wedge of the B system retreats to the exterior of the horizon, and the entanglement wedge of the auxiliary R system is the shaded blue island behind the horizon.

For the simple model described above, we will derive this answer using the replica trick, along with corrections. We will start by deriving a qualitative first approximation to the full answer, and then gradually progress to calculating the full non-perturbative answer in section 2.5.

2.2 Computation of the purity

We would like to use the 2d gravity path integral to compute the entropy of system R, in the state $|\Psi\rangle$ given in (2.5). We can start by discussing the density matrix ρ_{R}:

$$\rho_{\mathsf{R}} = \frac{1}{k} \sum_{i,j=1}^{k} |j\rangle\langle i|_{\mathsf{R}} \ \langle\psi_i|\psi_j\rangle_{\mathsf{B}}. \tag{2.8}$$

The matrix elements of ρ_{R} are gravity amplitudes $\langle\psi_i|\psi_j\rangle$. These are determined (up to normalization of the states $|\psi_i\rangle$) by a gravity calculation with the following boundary conditions

$$\langle\psi_i|\psi_j\rangle = \quad \text{\small}i \quad \underset{\text{\small}j}{\raisebox{0pt}{}} \tag{2.9}$$

Here, the black line is the asymptotic boundary of type (2.3), with renormalized length β. We drew the line with a wiggle to emphasize that the boundary conditions do not constraint its shape. The arrow represents the direction of time evolution, from the ket to the bra.[10] At the locations where the dashed blue lines intersect the solid black line, we require that an EOW brane of type i or j should intersect the asymptotic boundary. Here and below, dashed lines carry the index of the EOW brane state.

The leading gravity configuration that satisfies these boundary conditions is the following classical solution

$$\langle\psi_i|\psi_j\rangle \approx \tag{2.10}$$

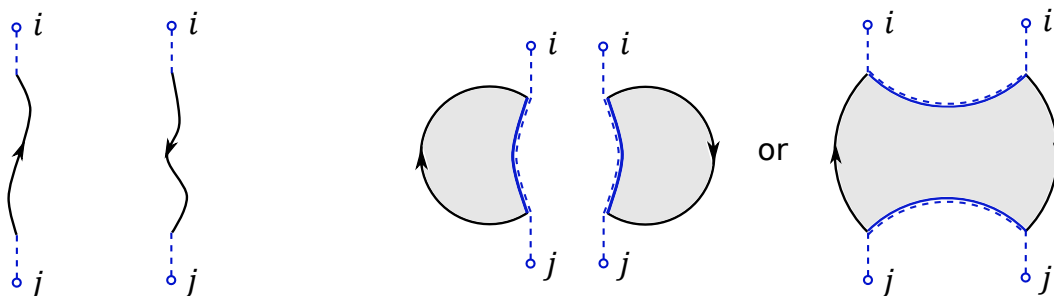[10]In a theory with time-reversal symmetry, we would not draw such an arrow.

Figure 3. The boundary conditions for $|\langle\psi_i|\psi_j\rangle|^2$ are shown at left, and two ways of filling in the geometry are shown at right. To compute the purity, we want to sum over i, j by connecting the dashed lines together. For the disconnected geometry, this will lead to a single k index loop, and for the connected geometry it will lead to two loops.

The black asymptotic boundary (with arrow) borders a portion of the hyperbolic disk, and a solid blue EOW brane follows a geodesic between the i and j endpoints. An important feature is that the same EOW brane connects to both the i index and the j index. Because we assume that these correspond to orthogonal internal states of the EOW brane, the result for $\langle\psi_i|\psi_j\rangle$ will be proportional to δ_{ij}. So, based on this computation, it looks like ρ_R will be maximally mixed, with entropy $\log(k)$:

$$\rho_\mathsf{R} \overset{?}{=} \frac{1}{k} \sum_{i=1}^{k} |i\rangle\langle i|. \tag{2.11}$$

However, we can also try to compute the entropy directly, using the replica trick. Let's start by considering the so-called "purity" $\mathrm{Tr}(\rho_\mathsf{R}^2)$, which is closely related to the Renyi 2-entropy. From (2.8), one can easily work out that

$$\mathrm{Tr}(\rho_\mathsf{R}^2) = \frac{1}{k^2} \sum_{i,j=1}^{k} |\langle\psi_i|\psi_j\rangle|^2. \tag{2.12}$$

The boundary conditions to compute $|\langle\psi_i|\psi_j\rangle|^2$ are shown in the left panel of figure 3. To compute the purity, we will sum over i, j by connecting the dashed lines in the obvious way.

The crucial point is that there are two different ways of filling these boundary conditions in with 2d geometry, as shown in the right panel of figure 3. We can either have a disconnected geometry with the topology of two disks, or a connected "Euclidean wormhole" geometry with the topology of a single disk.[11] To study a two-sided version of this model where we don't take the $\mathbb{Z}_2$ quotient, we would glue two copies of the EOW brane geometries together along the EOW branes. Then the topology would be two disks in the disconnected case, and one cylinder in the connected case.

In both topological classes, there is a classical solution in JT gravity. This is somewhat nontrivial for the connected (wormhole) case, for the following reason. There is a parameter

---

[11]Euclidean wormholes are sometimes also referred to as spacetime wormholes, to differentiate them from the more commonly studied *spatial* wormholes such as Einstein-Rosen bridges.

that characterizes the wormhole geometry, which we can take to be the regularized distance from one of the asymptotic boundaries to the other, through the wormhole. By itself, the action of JT gravity provides a pressure that would prefer this distance to be large. But the tension provided by the mass μ of the particle or EOW brane prefers the length to be short, and provides a counterbalancing force that leads to a stable minimum for the action.[12]

To describe the contributions of these geometries to $\text{Tr}(\rho_\mathsf{R}^2)$, we will use the notation $Z_n = Z_n(\beta)$ to represent the gravity path integral on a disk topology with a boundary that consists of alternating segments of n physical boundaries of renormalized lengths β, and n EOW branes. Using this notation, we can evaluate the sum of the two contributions in figure 3 as

$$\text{Tr}(\rho_\mathsf{R}^2) = \frac{kZ_1^2 + k^2 Z_2}{(kZ_1)^2} = \frac{1}{k} + \frac{Z_2}{Z_1^2}. \tag{2.13}$$

In the numerator, we have the contributions of the two geometries in figure 3: the disconnected geometry at left has one k-index loop, and two copies of the geometry that defines Z_1. The connected geometry at right has two k-index loops, and a single copy of the geometry that defines Z_2. In the denominator, we have divided by the gravity computation that normalizes the density matrix.[13]

We will work out exact formulas for Z_n in JT gravity below, but the basic point can be seen already in a very crude approximation, where we retain only the dependence on the topological S_0 term in the JT gravity action (2.2). This term weights the contribution of a given topology by $e^{S_0 \chi}$, where χ is the Euler characteristic. Since the topology relevant for Z_n is disk-like for any value of n, and since $\chi = 1$ for the disk, we will have

$$Z_n \propto e^{S_0}. \tag{2.14}$$

Using this formula, we see that the second term on the r.h.s. of (2.13) is proportional to e^{-S_0}. So

$$\text{Tr}(\rho_\mathsf{R}^2) = k^{-1} + e^{-S_0} \qquad \text{(schematic).} \tag{2.15}$$

If k is reasonably small the disconnected geometry dominates, and we will find that the purity is $1/k$, which is consistent with (2.11). However, if k gets very big, then the connected geometry dominates and we will find the purity is e^{-S_0}, independent of k. This interchange of dominance of the two saddlepoints is the basic mechanism that prevents the entropy (here the Renyi entropy) of the radiation from growing indefinitely. Note that because the state of the entire system is pure, this implies that the entropy of the excitations in the black hole (EOW brane states) remains finite even when k is very large. The basic mechanism for this is the small nonorthogonality of these states, as we discuss in the next section. When k becomes of order $e^{S_{\text{BH}}}$, this nonorthogonality adds up to a large effect.

---

[12]Note, however, that having a classical solution isn't necessary for the contribution to be meaningful.

[13]There is a subtlety here. If we normalize the state so that $\text{Tr}(\rho_\mathsf{R}) = 1$, we will find that $\text{Tr}(\rho_\mathsf{R})^2$ is not exactly one. This can be interpreted as due to the small fluctuation in the normalization of the states. This effect can be taken into account, but in the planar approximation that we will describe below, it is consistent to ignore it.

2.3 Factorization and averaging

The calculation in the previous section raises some important subtleties. For example, how can this result for $\text{Tr}(\rho_R^2)$ be reconciled with the formula for ρ_R in (2.11)? To address this, let's first consider the amplitude $\langle \psi_i | \psi_j \rangle$. For this quantity, the gravity path integral gives two seemingly-contradictory answers

$$\langle \psi_i | \psi_j \rangle = \delta_{ij}, \qquad |\langle \psi_i | \psi_j \rangle|^2 = \delta_{ij} + \frac{Z_2}{Z_1^2}. \qquad (2.16)$$

The first of these equations follows from the gravitational calculation in (2.10): the answer is proportional to δ_{ij} because the same EOW brane connects to both the i index and the j index. The two terms in the second equation correspond to the two terms in the r.h.s. of figure 3. Note that in the second term i does not have to be equal to j.

Of course, the two equations in (2.16) are incompatible in a strict sense. However, suppose that we imagine that the true quantum amplitude is

$$\langle \psi_i | \psi_j \rangle = \delta_{ij} + e^{-S_0/2} R_{ij} \qquad (2.17)$$

where R_{ij} is a random variable with mean zero. Then if we interpret the gravity path integral as computing some type of average over the microscopic R_{ij} quantities, then we would interpret the gravity answers as telling us that

$$\overline{\langle \psi_i | \psi_j \rangle} = \delta_{ij}, \qquad \overline{|\langle \psi_i | \psi_j \rangle|^2} = \delta_{ij} + \frac{Z_2}{Z_1^2}. \qquad (2.18)$$

Here, the bar indicates an average over R_{ij}. Now the equations are compatible, and the extra term in the second equation tells us the variance of R_{ij}. Note that the correction to orthogonality in (2.17) is very small. But the large number of terms in (2.12) enables these small corrections to dominate the final answer for $k \gg e^{S_0}$, giving a qualitatively different result than (2.11).

The conclusion seems to be that the gravity path integral is not literally computing quantum amplitudes, but is instead computing some coarse-grained version, where we average over some microscopic information R_{ij}. The lack of factorization of the resulting quantities is familiar from the connection between Euclidean wormholes and disorder.[14] This connection goes back to the work of Coleman [54] and of Giddings and Strominger [55], and has been recently discussed in JT gravity in [25]. It raises important questions in the present context, which we will return to in section 7.

In appendix D we give a precise description of the random ensemble that is dual to the JT gravity plus EOW branes model.

2.4 First look at the von Neumann entropy

The standard way to compute the von Neumann entropy using replicas is to use

$$S_R = -\text{Tr}(\rho_R \log \rho_R) = -\lim_{n \to 1} \frac{1}{n-1} \log \text{Tr}(\rho_R^n). \qquad (2.19)$$

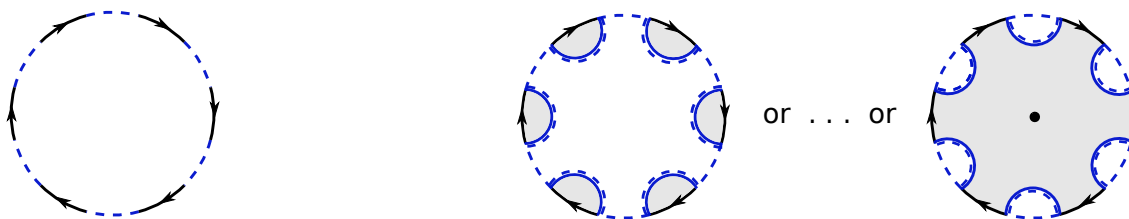[14]A different type of factorization problem in JT gravity has been discussed in [51–53].

Figure 4. The boundary conditions for $\text{Tr}(\rho_R^n)$ with $n = 6$ are shown at left, along with two extreme ways of filling in the geometry, corresponding to the completely connected and completely disconnected options. Note that the geometry at right contains a fixed point of the $\mathbb{Z}_n$ symmetry that rotates the replicas.

For this to be useful, it is important to be able to compute Renyi entropies in a way that has a good continuation in n. This is rather nontrivial, since as n grows large, there are many different geometries that can fill in the boundary conditions for computing $\text{Tr}(\rho^n)$ (shown in the left panel of figure 4). We will show how to handle this problem in the next section.

First, though, we can do a simplified analysis that is appropriate in two extreme limits. In the case where $k \ll e^{S_{\text{BH}}}$, completely disconnected geometries with the topology of n disks dominate. They have a single k-index loop, and give the contribution

$$\text{Tr}(\rho_R^n) \supset \frac{kZ_1^n}{k^n Z_1^n} = \frac{1}{k^{n-1}}. \tag{2.20}$$

Using (2.20), we find $S_R = \log(k)$, consistent with (2.7).

In the opposite limit where $k \gg e^{S_{\text{BH}}}$, the completely connected geometry with the topology of a single disk dominates. There are n index loops (see figure 4), and the contribution is

$$\text{Tr}(\rho_R^n) \supset \frac{k^n Z_n}{k^n Z_1^n} = \frac{Z_n}{Z_1^n}. \tag{2.21}$$

The k factors cancel out, so this answer is purely gravitational. In order to compute the von Neumann entropy, we need to continue Z_n to near $n = 1$. A trick for doing this (see section 3.2 of [7]) is to notice that the geometry associated to Z_n has a $\mathbb{Z}_n$ replica symmetry, and that the $\mathbb{Z}_n$ quotient of the geometry can be continued in n. In the limit $n \to 1$, this becomes the original unreplicated geometry, and the computation of the von Neumann entropy reduces to $S_0 + 2\pi\phi_h$, where ϕ_h is the value of the dilaton at the horizon, which is the fixed point of the $\mathbb{Z}_n$ symmetry in the $n \to 1$ limit. So we conclude that in this phase, the answer is just the thermodynamic entropy of the black hole. Again, this is consistent with (2.7).

The main conceptual point that we want to emphasize is that the island extremal surface descends from a Renyi entropy computation that involves replica wormholes.

2.5 Summation of planar geometries

So far, we have discussed the completely disconnected and completely connected replica geometries. In the regimes where these dominate, we get the two different answers for the von Neumann entropy, predicted by the island conjecture. In order to make this more precise, and to understand the transition between the two regimes, we need to sum over
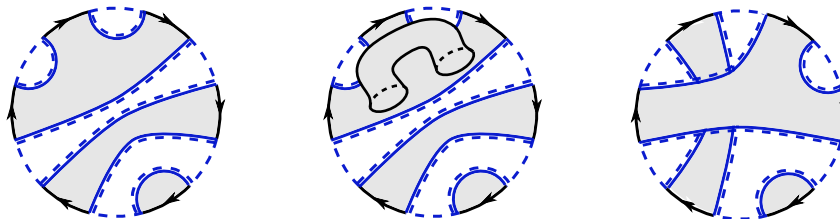
Figure 5. The left figure is an example of planar geometry that we need to include in our analysis. The middle figure has an extra handle and is down by e^{-2S_0}. The right figure involves a crossing, and is down by k^{-2} (it has two dashed index loops instead of four).

replica geometries that are intermediate between the completely disconnected and complete connected cases shown in figure 4.

The starting point is the boundary condition for $\mathrm{Tr}(\rho_R^n)$, shown in the left panel of figure 4. In principle, we need to sum over all ways of filling this boundary condition in. To simplify, we will assume that e^{S_0} and k are both large. In this approximation, we only need to sum over geometries that are planar, in the sense explained in figure 5. Non-planar geometries are suppressed either by powers of e^{-2S_0} (for adding handles), or by powers of $1/k^2$ (for introducing crossings).

Depending on the relative size of k or e^{S_0}, either highly disconnected or highly connected geometries will tend to dominate. We do not assume any particular relationship between k and e^{S_0}, so it will be necessary to sum over all planar geometries. In order to do this, it is convenient to define the resolvent matrix $R_{ij}(\lambda)$ of ρ_R:

$$R_{ij}(\lambda) = \left(\frac{1}{\lambda\mathbb{1} - \rho_R}\right)_{ij} = \frac{1}{\lambda}\delta_{ij} + \sum_{n=1}^{\infty}\frac{1}{\lambda^{n+1}}(\rho_R^n)_{ij}. \tag{2.22}$$

We will write down a Schwinger-Dyson equation for R_{ij} using the planarity property and then use its solution to find the entanglement spectrum of ρ_R.[15]

The boundary conditions for R_{ij} correspond to an infinite sum over different numbers of asymptotic AdS$_2$ boundaries, connected by the k index lines associated to the R system:

$$\tag{2.23}$$

The dashed index lines come with factors of $1/\lambda$, and the solid lines with arrows come with factors $1/(kZ_1)$ that normalize the gravitational path integral. We fill these in with all possible planar geometries:

$$\tag{2.24}$$

---

[15]This analysis was inspired by the "free probability" results discussed in [56] and e.g. figure 1 of the earlier [57].

For the one-boundary terms in (2.23), there is only a single bulk geometry that can fill it in. For the two-boundary term, there are two possible geometries, and we sum over them in (2.24). For the three-boundary term (not shown), there would be five possible geometries.

We can write a Schwinger-Dyson equation that sums these geometries:

$$(2.25)$$

On the r.h.s., the second term sums all planar geometries in which the first boundary is disconnected from all other boundaries. The third term sums all planar geometries in which the first boundary is connected to one other boundary, and so on.

To write this identity as an equation, it is convenient to define R as the trace of the resolvent matrix,

$$R(\lambda) = \sum_{i=1}^{k} R_{ii}(\lambda). \tag{2.26}$$

Then (2.25) is equivalent to

$$R_{ij}(\lambda) = \frac{1}{\lambda}\delta_{ij} + \frac{1}{\lambda}\sum_{n=1}^{\infty} \frac{Z_n}{(kZ_1)^n} R(\lambda)^{n-1} R_{ij}(\lambda). \tag{2.27}$$

We can think of the δ_{ij}/λ as a "bare propagator". It is the first term in (2.24). In the remaining terms, n labels the number of boundaries of the geometry that contains the leftmost boundary. The Z_n factor is the gravitational path integral of an n boundary geometry, which we will analyze below. For each of the n boundaries, we divided by kZ_1 in order to normalize the density matrix ρ_R. Taking the trace of (2.27), we find

$$\lambda R(\lambda) = k + \sum_{n=1}^{\infty} Z_n \frac{R(\lambda)^n}{k^n Z_1^n}. \tag{2.28}$$

In JT gravity, it is possible to make this equation very explicit.[16] The exact formula for Z_n can be worked out using the boundary particle formalism [58, 59]. We decompose the path integral for Z_n in the following way

$$(2.29)$$

---

[16]For general two dimensional dilaton gravity theories, one can do the summation at the classical level and for large μ. See appendix B.

Here, the object $I_{2n}(\ell_1, \ldots, \ell_{2n})$ is the JT gravity path integral with $2n$ geodesic boundaries of fixed regularized lengths ℓ. It has the following expression [58]

$$I_{2n}(\ell_1, \ldots, \ell_{2n}) = 2^{2n} \int_0^\infty ds\, \rho(s) K_{2is}\left(4e^{-\frac{\ell_1}{2}}\right) \ldots K_{2is}\left(4e^{-\frac{\ell_{2n}}{2}}\right); \quad \rho(s) = \frac{s}{2\pi^2} \sinh(2\pi s), \tag{2.30}$$

where K is the modified Bessel function. The object φ in (2.29) is the Hartle-Hawking state in the geodesic basis [51]. It computes the path integral from the asymptotic boundary (characterized by renormalized length β) to the geodesic of regularized length ℓ. It has the expression [58, 59]

$$\varphi_\beta(\ell) = 4e^{-\frac{\ell}{2}} \int_0^\infty ds\, \rho(s) e^{\frac{-\beta s^2}{2}} K_{2is}\left(4e^{-\frac{\ell}{2}}\right). \tag{2.31}$$

The boundary of I_{2n} consists of n geodesics that correspond to EOW branes, and n geodesics that need to be glued to Hartle-Hawking states. In both cases, the procedure is similar: we integrate over the length of the geodesic, with a "wave function" that is either the $e^{-\mu\ell}$ weighting for the EOW brane, or the $\varphi_\beta(\ell)$ weighting for the Hartle-Hawking states. Including the correct measure factor for the integral over geodesic lengths (see appendix A), the formula for Z_n is

$$\begin{aligned} Z_n &= e^{S_0} \int_0^\infty d\ell_1 \ldots d\ell_{2n} e^{\frac{\ell_1 + \ldots + \ell_{2n}}{2}} I_{2n}(\ell_1, \ldots \ell_{2n}) \varphi_\beta(\ell_1) e^{-\mu\ell_2} \ldots \varphi_\beta(\ell_{2n-1}) e^{-\mu\ell_{2n}} \\ &= e^{S_0} \int_0^\infty ds\, \rho(s) y(s)^n; \qquad y(s) = e^{-\frac{\beta s^2}{2}} 2^{1-2\mu} |\Gamma\left(\mu - \frac{1}{2} + is\right)|^2. \end{aligned} \tag{2.32}$$

In going to the second line from the first one, we inserted the integral representations (2.30) and (2.31), and then did the integrals using formulas discussed in appendix A. In $y(s)$, the Boltzmann factor $\exp(-\frac{\beta s^2}{2})$ and the Gamma function factor $2^{1-2\mu}|\Gamma(\mu - \frac{1}{2} + is)|^2$ come from the integral with the Hartle-Hawking state and with the brane state, respectively.

Physically, the s parameter can be viewed as giving the energy $s^2/2$ of a particular asymptotic region. The main simplification in the above calculation is the fact that the energy must be the same for all asymptotic boundaries in a single connected geometry.

Mathematically, (2.32) gives an integral representation for Z_n, with the property that the n dependence is very simple inside the integral. The sum needed for the Schwinger-Dyson equation (2.25) becomes a geometric series. In order to unclutter the equations that follow, we will define rescaled variables

$$\boldsymbol{\rho}(s) = e^{S_0}\rho(s), \qquad w(s) = \frac{y(s)}{Z_1}. \tag{2.33}$$

Then, after summing the geometric series, one finds that the equation for the resolvent is

$$\lambda R = k + \int_0^\infty ds\, \boldsymbol{\rho}(s) \frac{w(s)R}{k - w(s)R}. \tag{2.34}$$

This equation contains all of the information about the entanglement spectrum of ρ_R, and it is exact in the planar approximation.[17] By solving the equation for $R(\lambda)$, and then taking

---

[17]A similar equation applies to arbitrary dilaton gravity theories, at least in the classical limit and with large μ, see equation (B.5).

the discontinuity across the real axis, one can find the density of eigenvalues of the matrix ρ_R:

$$D(\lambda) = \frac{1}{2\pi i}[R(\lambda - i\epsilon) - R(\lambda + i\epsilon)]. \tag{2.35}$$

The von Neumann entropy can then be computed by evaluating

$$S(\mathsf{R}) = -\int d\lambda \, D(\lambda) \, \lambda \log(\lambda). \tag{2.36}$$

2.5.1 Microcanonical ensemble: Page's result

Let's first use this planar resummation to compute the entanglement spectrum for the case where the black hole is in a microcanonical ensemble. This means that we fix the energy in each asymptotic region, rather than fixing the renormalized length β. Let s be the chosen value of $s = \sqrt{2E}$, and let Δs be the width of a small interval that defines the microcanonical ensemble. It is convenient to define the boldface quantities $\mathbf{S}$, $\mathbf{Z_n}$ and $\mathbf{w}$:

$$e^{\mathbf{S}} = \rho(s)\Delta s, \qquad \mathbf{Z}_n = \rho(s)y(s)^n\Delta s, \qquad \mathbf{w}(s) = \frac{y(s)}{\mathbf{Z}_1} = e^{-\mathbf{S}}. \tag{2.37}$$

Here $\mathbf{S}$ is the entropy of our microcanonical ensemble, and $\mathbf{Z}_n$ is the microcanonical version of Z_n.

The advantage of this microcanonical ensemble is that the resolvent equation (2.28) simplifies to a quadratic equation for R. This equation, and the corresponding solution for the density of eigenvalues (2.35), are given by

$$R(\lambda)^2 + \left(\frac{e^{\mathbf{S}} - k}{\lambda} - ke^{\mathbf{S}}\right)R(\lambda) + \frac{k^2 e^{\mathbf{S}}}{\lambda} = 0;$$

$$D(\lambda) = \frac{ke^{\mathbf{S}}}{2\pi\lambda}\sqrt{\left[\lambda - \left(k^{-\frac{1}{2}} - e^{-\frac{\mathbf{S}}{2}}\right)^2\right]\left[\left(k^{-\frac{1}{2}} + e^{-\frac{\mathbf{S}}{2}}\right)^2 - \lambda\right]} + \delta(\lambda)(k - e^{\mathbf{S}})\,\theta(k - e^{\mathbf{S}}). \tag{2.38}$$

These equations are precisely the same ones found by Page for the entanglement spectrum of a subsystem of dimension k in a random state of total dimension $ke^{\mathbf{S}}$, in the planar approximation [2]. The fact that we get the random state answer can be understood from the fact that for fixed energy, the random ensemble dual to our JT gravity + EOW branes model (see appendix D) is the same random state ensemble discussed by Page.

1. When $k \ll e^{\mathbf{S}}$, the range of the spectrum is very narrow and the first term in $D(\lambda)$ is a very narrow semicircle, roughly a delta function. So the spectrum consists of k eigenvalues of size $1/k$. This describes a maximally mixed density matrix and the von Neumann entropy is $\log k$.

2. As k approaches $e^{\mathbf{S}}$ (the Page time), the distribution develops a $1/\sqrt{\lambda}$ singularity at the origin. After the Page time, the smooth part of the distribution contains only $e^{\mathbf{S}}$ eigenvalues. The remaining $k - e^{\mathbf{S}}$ are exactly zero. The von Neumann entropy has a rather sharp transition from $\log k$ to $\mathbf{S}$ during the Page transition: $S = \log m - \frac{m}{2n}$, where $m = \min(k, e^{\mathbf{S}})$, and $n = \max(k, e^{\mathbf{S}})$ [2].

3. When $k \gg e^{\mathbf{S}}$, the smooth term can again be approximated as a delta function. The distribution describes a density matrix that is maximally mixed on an $e^{\mathbf{S}}$ dimensional subspace, and unentangled in the rest. The von Neumann entropy stays equal to $\mathbf{S}$.
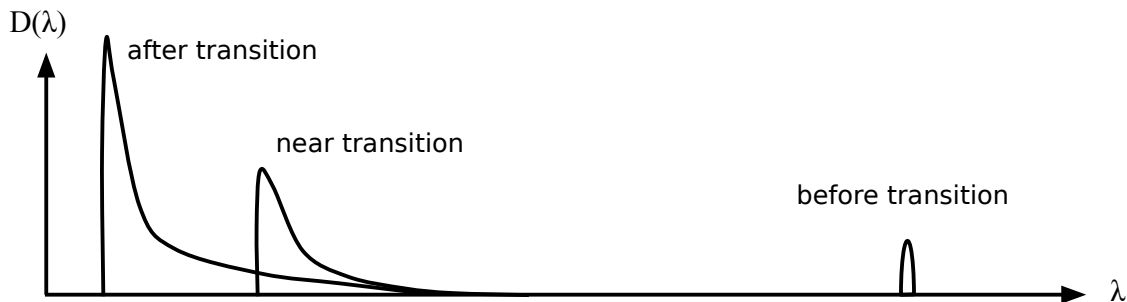
Figure 6. Sketch (not to scale) of the entanglement spectrum in the canonical ensemble. Before the Page transition, the density of states has a very narrow distribution. Near the Page transition, it becomes a shifted thermal spectrum (2.51), with a cutoff at some energy. After the Page time, it becomes an ordinary thermal spectrum, again with a cutoff at some energy.

2.5.2 Canonical ensemble: smoothing out the transition

In the canonical ensemble, we did not find a way to solve the resolvent equation (2.34) exactly. However, we can get some intuition by solving the equation numerically. To do this, one can first evaluate λ as a function of R, find the locus where λ is real, and then compute the inverse. The results of this procedure are sketched in figure 6. Long before the Page transition, the distribution of eigenvalues $D(\lambda)$ is very narrow, localized near $\lambda = 1/k$ as in the microcanonical case. Long after the Page transition, the distribution resembles the thermal spectrum of the black hole, with a rather sharp cutoff at $s = s_k$ chosen so that the total number of eigenvalues is k:

$$k = \int_0^{s_k} \mathrm{d}s \, \rho(s), \qquad \text{(definition of } s_k\text{)}. \tag{2.39}$$

In between the two regimes, the curve smoothly interpolates, with no singular feature during the transition.

We will now try to analyze the resolvent equation approximately, with the goal of computing the von Neumann entropy in the semiclassical, small G_N regime. In our formulas for JT gravity, we did not include a G_N parameter explicitly, but it can be restored by taking

$$\beta \to G_N \beta, \tag{2.40}$$

so the semiclassical small G_N limit corresponds to small β. We give a much more detailed analysis of the same problem in appendix F.

As a first step, we can determine the location of the bottom edge in the distribution $D(\lambda)$, i.e. the smallest eigenvalue λ_0. This corresponds to a branch point in $R(\lambda)$, which means a location where $d\lambda/dR = 0$. To determine this point, we first write the resolvent equation in the form

$$\lambda = \frac{k}{R} + \int_0^\infty \mathrm{d}s \, \rho(s) \frac{w(s)}{k - w(s)R}. \tag{2.41}$$

In the region near the lower endpoint of the spectrum, R is real and negative, and the above integral can be approximated by dividing the s integral into two intervals where the

two different terms in the denominator dominate:

$$\lambda \approx \frac{k}{R} - \frac{1}{R} \int_0^{s_R} \mathrm{d}s \, \rho(s) + \frac{1}{k} \int_{s_R}^\infty \mathrm{d}s \, \rho(s) w(s). \tag{2.42}$$

Here s_R is defined by $w(s_R)R = -k$. This approximation is justified in more detail in appendix F.

Setting to zero the derivative of (2.42) with respect to R, we find that the first two terms on the r.h.s. should be equal. Comparison with (2.39) then implies that $s_R \approx s_k$. Plugging this back into (2.42), we find that the smallest eigenvalue of ρ_R, and the corresponding value of R are

$$\lambda_0 \approx \frac{1}{k} \int_{s_k}^\infty \mathrm{d}s \, \rho(s) w(s), \qquad R(\lambda_0) \approx -\frac{k}{w(s_k)}. \tag{2.43}$$

The next step is to write the equation for the resolvent (2.34) and break up the s integral into two parts:

$$\lambda R \approx k + \int_0^{s_k} \mathrm{d}s \, \rho(s) \frac{w(s)R}{k - w(s)R} + \frac{R}{k} \int_{s_k}^\infty \mathrm{d}s \, \rho(s) w(s) \tag{2.44}$$

In the region from s_k to infinity, we replaced the factor of $k - w(s)R$ in the denominator by k. This is justified on the following grounds. First, we are going to study the equation in the region $\lambda > \lambda_0$, and in this region $|R| < |R(\lambda_0)|$. Second, $w(s)$ is a decreasing function of s, so for $s > s_k$, we have $|w(s)R| \leq |w(s_k)R(\lambda_0)| = k$. Intuitively, we have separated the s integral into two terms in (2.44) corresponding to the "pre-Page" and "post-Page" parts of the thermal ensemble. The high energy states with $s > s_k$ are effectively before the Page transition, and the planar resummation (represented by the nontrivial denominator) is not necessary for these states.

A nice feature of (2.44) is that the final term can be recognized as $\lambda_0 R$, so (2.44) can be rewritten

$$(\lambda - \lambda_0)R \approx k + \int_0^{s_k} \rho(s) \frac{w(s)R}{k - w(s)R}. \tag{2.45}$$

We can solve this equation in an approximation where the second term is a small perturbation, and the zero-th order solution is just $k/(\lambda - \lambda_0)$. Iterating the equation once, we find the first-order solution

$$R(\lambda) \approx \frac{k}{\lambda - \lambda_0} + \frac{1}{\lambda - \lambda_0} \int_0^{s_k} \mathrm{d}s \, \rho(s) \frac{w(s)}{\lambda - \lambda_0 - w(s)}. \tag{2.46}$$

This approximation is good as long as $\int_0^{s_k} \mathrm{d}s \, \rho(s) \frac{w(s)}{\lambda - \lambda_0 - w(s)} \ll k$, which will be true as long as

$$\lambda > \lambda_0 + w(s_k - \delta), \tag{2.47}$$

with δ a control parameter. So we conclude that for eigenvalues λ satisfying (2.47), we can compute the resolvent to good accuracy. Unfortunately, there is a gap between this value and the actual bottom of the spectrum λ_0, and we do not have control over the distribution of eigenvalues in this region. Before and during the Page transition, we are saved by the fact that $\lambda_0 \gg w(s_k)$, so for the purposes of computing the von Neumann entropy, this is

a narrow region where the density of eigenvalues can be approximated by a delta function with unknown weight. After the Page transition, this is no longer true, but the eigenvalues in the unknown region contribute a small amount to the entropy.

So, using (2.44) in the region (2.47) and parametrizing our ignorance in the remaining eigenvalues with a delta function, we have

$$D(\lambda) \approx \#\delta(\lambda - \lambda_0) + \int_0^{s_k - \delta} \mathrm{d}s\, \rho(s)\delta(\lambda - \lambda_0 - w(s)). \tag{2.48}$$

In the region between $s_k - \delta$ and s_k, $\lambda_0 \gg w(s)$, so we can approximate $\lambda_0 + w(s) \approx \lambda_0$. This means that we can let the integral run all the way to s_k, at the cost of changing the coefficient of the delta function piece. The resulting combined coefficient can be shown to vanish, by the following argument. The distribution of eigenvalues has to satisfy two normalization conditions,

$$\int \mathrm{d}\lambda\, D(\lambda) = k, \qquad \int \mathrm{d}\lambda\, D(\lambda)\lambda = 1. \tag{2.49}$$

The first condition says that ρ_R has k total eigenvalues, and the second condition says that ρ_R is normalized. The first of these equations implies that

$$D(\lambda) = \int_0^{s_k} \mathrm{d}s\, \rho(s)\delta(\lambda - \lambda_0 - w(s)) \tag{2.50}$$

is correctly normalized without any further delta function piece. One can check that the second condition follows from our result for λ_0 in (2.43).

This is our final expression for the entanglement spectrum. It can be given a very simple interpretation: we have the spectrum of the first k states of the thermal spectrum of the black hole, shifted so that the total normalization is one. This shift can be understood as the contribution of all of the remaining states in the spectrum of the black hole:

$$\rho_\mathsf{R} \simeq \mathbb{P}\rho_{\mathrm{BH}}\mathbb{P} + \frac{\mathbb{1}}{k}\mathrm{Tr}\!\left(\bar{\mathbb{P}}\rho_{\mathrm{BH}}\bar{\mathbb{P}}\right). \tag{2.51}$$

Here $\mathbb{P}$ is the projector into the post-Page subspace with $s < s_k$, and $\bar{\mathbb{P}}$ is the projector into the pre-Page subspace with $s > s_k$. The operator ρ_{BH} is the density matrix of the thermal ensemble (including the brane) for the black hole.

Using this result, one can evaluate the von Neumann entropy. Inserting (2.50) into (2.36) and doing the λ integral using the delta function, we find

$$S \approx -\int_0^{s_k} \mathrm{d}s\, \rho(s)(\lambda_0 + w(s))\log(\lambda_0 + w(s)). \tag{2.52}$$

A careful analysis shows that the error in this approximation peaks around the Page transition, when it is of order G_N (see appendix F for details). As a function of k, this expression exhibits a smoothed-out version of the Page transition. It differs from the naive answer $\min(k, S_{\mathrm{BH}})$ by an amount of order $G_N^{-1/2}$ near the Page transition, but one can show that it agrees to within $O(1)$ precision with the average of the micro-canonical answer over the full thermal ensemble. In figure 7 we plot the exact formula for the
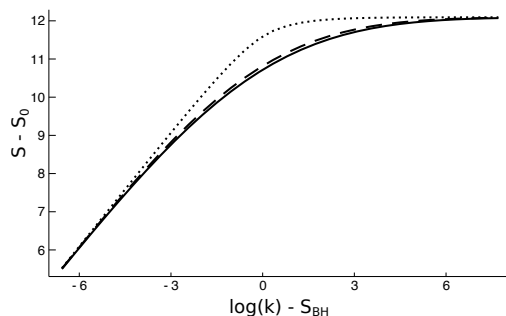
Figure 7. We plot the exact entropy in the planar approximation (solid) and the approximation (2.52) (dashed) for the case $\beta = 3$ and large μ. The answer in the microcanonical ensemble with the same entropy is also shown (dotted). The difference between the curves is largest near the Page transition. For small β, the transition in the canonical curves gets smoothed out over a large region of size $\beta^{-1/2} \sim G_N^{-1/2}$ in $\log(k)$, while the transition in the microcanonical curve always looks the same. The maximum error between the canonical curve and the simple approximation (2.52) will be of order $\beta \sim G_N$.

entropy for $\beta = 3$, obtained by solving the resolvent equation numerically. We also plot the approximation (2.52) and the exact microcanonical answer for comparison.

We would like to emphasize one point about the continuation in n. Because they are sensitive to different parts of the thermal spectrum, the different Renyi entropies $\text{Tr}(\rho^n)$ experience Page transitions at different values of k. In fact, there are values of k for which the Page transition for all of the integer Renyi entropies has already taken place, but the von Neumann entropy is still very close to $\log(k)$. In this regime, a fully connected geometry dominates the computation of the Renyi entropy for every integer n, but its continuation to $n = 1$ would give the wrong answer in the von Neumann limit. Nevertheless, the full resummation gives the right answer.

3 Reconstruction behind the horizon via the Petz map

In the previous section, we showed (in the simple model) that for large k, the entanglement entropy saturates at the thermodynamic entropy of the black hole, and the RT surface is at the horizon. This means that the entanglement wedge of the radiation R contains an island behind the horizon.

The location of the entanglement wedge is significant because of the notion of "entanglement wedge reconstruction,"[18] which implies that in the large k phase, the radiation system R should describe the island. The arguments in favor of entanglement-wedge reconstruction [8, 37–39] are strong but non-constructive. In the next section we will show directly from the gravitational path integral that operators acting in the R system can control the region behind the horizon. We will see that the Euclidean wormholes described above play an essential role, connecting the region behind the horizon to the R system.

---

[18]This was originally conjectured in [6, 35, 36], and established in [37–39] using the ideas of [8].

Although we will focus on the simple model in our discussion below, a similar argument applies to the general case of operator reconstruction within the entanglement wedge, e.g. the famous case of two intervals in the AdS$_3$ vacuum. The argument is essentially the same and we describe this generalization in appendix C.

3.1 Setup

We would like to slightly enrich the simple model discussed above, so that there is something nontrivial to reconstruct. To do this, we will imagine coupling our JT gravity system to a bulk field theory, containing some propagating degrees of freedom. For simplicity, we will assume that these propagating degrees of freedom cannot feel the difference between different states of the EOW brane.

We will consider a "code subspace" of states corresponding to small excitations of the bulk fields propagating on the background described so far (i.e. by the background corresponding to the state (2.5), and to the geometry in figure 1). A basis for the code subspace is provided by the states
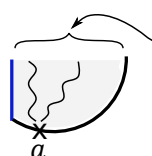
$$|\Psi_a\rangle = k^{-1/2} \sum_{j=1}^{k} |\psi_{aj}\rangle_{\mathsf{B}} |j\rangle_{\mathsf{R}}, \qquad a = 1, \ldots, d_{\mathsf{code}}. \tag{3.1}$$

Here, $|\psi_{aj}\rangle$ is a state in which the EOW brane is in state j, and we also have a small excitation of the propagating bulk fields, labeled by a.

Because of the variety of geometries that will be important below, we actually have to be somewhat more precise at this point, and define $|\psi_{aj}\rangle$ by a boundary condition rather than a bulk statement. What we really mean is that the a index corresponds to some insertion in the boundary conditions for the Euclidean path integral that defines the bulk state. So, for example, the boundary conditions for an inner product of two such states would be

$$\langle\psi_{ai}|\psi_{bj}\rangle = \begin{matrix} i & a & & j \\ \circ\text{-----}\times & \!\!\!\!\!\!\searrow\!\! & b & \text{-----}\circ \\ & & \times & \end{matrix} \tag{3.2}$$

where the labeled "×" marks represent the insertions associated to a and b. We assume that e.g. the a insertion is arranged so that on the leading disk topology, it produces a particular state of the bulk field theory $|a\rangle_{\mathsf{FT}}$,

$$\tag{3.3}$$

Relative to the state with no insertions, $|a\rangle_{\mathsf{FT}}$ could contain excitations behind the horizon, outside the horizon, or both. To simplify the formulas below, we will assume that the bulk states are indexed and normalized so that to a good approximation $\langle a|b\rangle_{\mathsf{FT}} = \delta_{ab}$. We will specialize to a more concrete model in section 3.7.

We let $\mathcal{O}_{\mathsf{FT}}$ be some operator that acts within this subspace of bulk states. We will denote its matrix elements in the bulk field theory states as $\mathcal{O}_{ab}$:

$$\mathcal{O}_{ab} \equiv \langle a|\mathcal{O}_{\mathsf{FT}}|b\rangle_{\mathsf{FT}}. \tag{3.4}$$

We define $\mathcal{O}$ to be the representation of the bulk operator $\mathcal{O}_{\mathsf{FT}}$ in the full boundary system

$$\mathcal{O} = \sum_{a,b=1}^{d_{\mathrm{code}}} |\Psi_a\rangle\langle\Psi_b| \times \mathcal{O}_{ab}. \tag{3.5}$$

Note that in general, this "global" representation of the operator acts nontrivially on both R and B.

Now we can get to the point. Suppose that we are in the phase with $k \gg e^{S_{\mathrm{BH}}}$, so that the entanglement wedge of R includes the island behind the horizon. Then if $\mathcal{O}_{\mathsf{FT}}$ acts within the "island" behind the horizon, the general arguments for entanglement-wedge reconstruction suggest that we should be able to find an operator $\mathcal{O}_{\mathsf{R}}$, acting only on R, such that

$$\langle\Psi_a|\mathcal{O}_{\mathsf{R}}|\Psi_b\rangle \approx \langle\Psi_a|\mathcal{O}|\Psi_b\rangle. \tag{3.6}$$

What we would like to do is show by a direct bulk calculation that this is possible. The tool we will use is the so-called "Petz map," which essentially gives a guess for what operator $\mathcal{O}_{\mathsf{R}}$ to choose. We will do a bulk calculation to show that it actually works, demonstrating entanglement-wedge reconstruction explicitly.

3.2 Petz Lite

Before we discuss the Petz map itself, we will go through a simpler ("Petz Lite") version, which will be good enough for a certain limited class of states. The Petz Lite map works as follows. Given an operator $\mathcal{O}$ on the combined system BR, we can define a operator on the R system using the partial trace

$$\mathcal{O}_{\mathsf{R}} = c_0 \mathrm{Tr}_{\mathsf{B}}(\mathcal{O}). \tag{3.7}$$

Here, the constant c_0 should be chosen so that the identity operator maps to the identity operator. This seems like a very naive guess for an operator mapping, and in fact this Petz Lite version will not work for general states. But it does work in a special class of states, analogous to the "fixed area" states of Akers, Rath, Dong, Harlow and Marolf [60, 61].

In JT gravity, fixed area is replaced by fixed dilaton ϕ, and we choose to fix the value of ϕ at the horizon to be the value that it has in the classical solution. The advantage of doing this is that when ϕ is fixed, we don't impose the equation $R + 2 = 0$ at that point, and instead we allow any value of R, including a delta function singularity that corresponds to a conical excess. This freedom makes replica geometries very easy to construct [60, 61]: we just take n copies of the $n = 1$ geometry, and glue them together in such a way that the horizon is a conical singularity with total angle $2\pi n$.

To proceed, let's define boldface states $|\boldsymbol{\psi}_{ai}\rangle, |\boldsymbol{\Psi}_a\rangle$ to be fixed-dilaton versions of the states $|\psi_{ai}\rangle, |\Psi_a\rangle$ defined above. Making this modification in (3.5) and tracing over the B system as in (3.7), we find the candidate operator

$$\mathcal{O}_{\mathsf{R}} = \frac{c_0}{k} \sum_{i,j=1}^{k} |i\rangle\langle j|_{\mathsf{R}} \sum_{a,b=1}^{d_{\mathrm{code}}} \langle\boldsymbol{\psi}_{bj}|\boldsymbol{\psi}_{ai}\rangle_{\mathsf{B}} \; \mathcal{O}_{ab}. \tag{3.8}$$

How does this operator manage to affect the region behind the horizon? Note that although $\mathcal{O}_{\mathsf{R}}$ acts only on R, its explicit form depends in detail on amplitudes $\langle\boldsymbol{\psi}_{ai}|\boldsymbol{\psi}_{bj}\rangle$ in the black

hole theory B. Let's imagine that some quantum computer will apply $\mathcal{O}_R$ to the R system. Then in order to compute these amplitudes, the quantum computer will do computations that are equivalent to simulating system B. This introduces a second copy of B into the game.

The basic point seems to be that the physical copy of the black hole B can connect via a Euclidean wormhole to this second copy, providing a geometrical connection between the black hole and R. This connection is equivalent to inserting an operator behind the horion of the physical black hole.

Let's see how this works in a very simple case, where the code subspace consists of two states: the state $|0\rangle_{FT}$ with no bulk particles excited, and $|1\rangle_{FT}$ with a single particle behind the horizon. We consider the operator

$$\mathcal{O}_{FT} = |1\rangle\langle 0|_{FT} \tag{3.9}$$

which creates a particle. For this case, Petz Lite gives

$$\mathcal{O}_R = \frac{c_0}{k} \sum_{i,j=1}^{k} |i\rangle\langle j|_R \langle\boldsymbol{\psi}_{0j}|\boldsymbol{\psi}_{1i}\rangle. \tag{3.10}$$

Here $|\boldsymbol{\psi}_{0j}\rangle$ is a state with no particle, and $|\boldsymbol{\psi}_{1i}\rangle$ is a state with the particle present. If we were to evaluate the amplitudes $\langle\boldsymbol{\psi}_{0j}|\boldsymbol{\psi}_{1i}\rangle$ using a single replica, we would find zero because the bulk field theory states with and without a particle are orthogonal. So the amplitude would vanish, and we would conclude that $\mathcal{O}_R = 0$.

However, as discussed in section 2.3, the gravity path integral seems to require an interpretation as a disorder average, where the amplitude $\langle\boldsymbol{\psi}_{0j}|\boldsymbol{\psi}_{1i}\rangle$ is only zero "on average" in some ensemble implicit in the gravitational description. So we should postpone our concern that the operator vanishes, and see what it does when we compute a matrix element:

$$\langle\boldsymbol{\Psi}_a|\mathcal{O}_R|\boldsymbol{\Psi}_b\rangle = \frac{c_0}{k^2} \sum_{i,j=1}^{k} \langle\boldsymbol{\psi}_{ai}|\boldsymbol{\psi}_{bj}\rangle\langle\boldsymbol{\psi}_{0j}|\boldsymbol{\psi}_{1i}\rangle. \tag{3.11}$$

On the r.h.s. we have a product of two gravitational amplitudes. The first factor corresponds to the physical black hole, and the second factor corresponds to the simulated one in the quantum computer acting on R. In the large k phase, the largest contribution to the expression on the r.h.s. comes from a connected geometry, where the physical black hole is connected to the simulated one by a wormhole. This wormhole is particularly easy to describe for the fixed-dilaton states. We simply glue two copies of the one-replica geometry together along a cut behind the horizon. This corresponds to a geometry with a 4π angle at the horizon, as sketched in figure 8. From the perspective of the physical black hole, this wormhole corresponds to an insertion of an operator behind the horizon. It is an operator with a particle present in the ket vector (top half of the cut) and no particle present in the bra vector (bottom half of the cut). So the operator acts as a creation operator behind the horizon.

This is the basic mechanism that underlies entanglement-wedge reconstruction. However, in order to go beyond fixed-dilaton states, we will have to use the full Petz map, rather than the Petz Lite version discussed above.
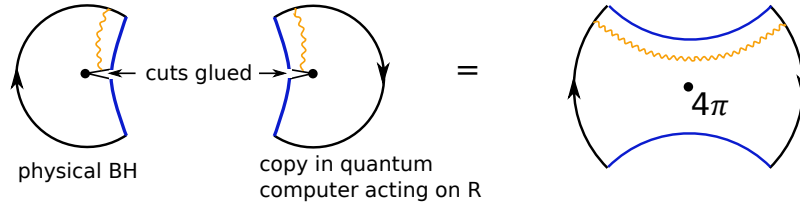
Figure 8. The two-replica geometry that contributes to the expectation value (3.11) for $a = 1, b = 0$ of the Petz Lite operator. The wavy line represents the bulk particle. Note that gluing the right geometry to the left geometry has the same effect (from the perspective of the physical left system) as inserting a creation operator behind the horizon. Another sketch of the glued geometry is shown at right. The marked point is the point of fixed dilaton, and it has conical angle 4π. The worldline of the particle created in the quantum computer (wiggly orange line) travels through the wormhole and into the physical black hole. The brane on the top half of each of the figures is a brane of type j and the brane on the bottom half is of type i.

3.3 Quantum information discussion of the Petz map

We now introduce the full Petz map, in its natural environment of quantum channels. The remarks below will not be necessary for our gravity discussion, but we include them for background.

A reversible quantum channel $\mathcal{N}$ can be defined by the property that it has a type of inverse $\mathcal{R}$, such that $(\mathcal{R} \circ \mathcal{N})(\rho) = \rho$ for all inputs to the channel $\mathcal{N}$. The Petz map [41, 42, 62] gives a universal formula for $\mathcal{R}$ in terms of $\mathcal{N}$, in cases where exact recovery is possible. We will not write the full formula, since we are only interested in a special case of it, but we refer the reader to [39, 43] for details and previous discussions in the context of entanglement wedge reconstruction.

As a representative example, let's consider a quantum channel that embeds $\mathcal{H}_{\mathsf{code}} \subseteq \mathcal{H}_A \otimes \mathcal{H}_{\bar{A}}$, and then traces over $\mathcal{H}_{\bar{A}}$. The channel is reversible if it is possible to recover $\mathcal{H}_{\mathsf{code}}$ from $\mathcal{H}_A$. In this situation, for any operator $\mathcal{O}_{\mathsf{code}}$ acting on $\mathcal{H}_{\mathsf{code}}$, the Petz map gives an equivalent operator $\mathcal{O}_A$ acting on A:

$$\mathcal{O}_A = \sigma_A^{-1/2} \mathrm{Tr}_{\bar{A}}(V \mathcal{O}_{\mathsf{code}} V^\dagger) \sigma_A^{-1/2}. \tag{3.12}$$

$$\sigma_A = \mathrm{Tr}_{\bar{A}}(\Pi_{\mathsf{code}}). \tag{3.13}$$

Here V is the encoding map from H_{code} into $\mathcal{H}_A \otimes \mathcal{H}_{\bar{A}}$, and Π_{code} is the projector onto the image of $\mathcal{H}_{\mathsf{code}}$ in $\mathcal{H}_A \otimes \mathcal{H}_{\bar{A}}$. By saying that $\mathcal{O}_A$ is equivalent to $\mathcal{O}_{\mathsf{code}}$, we mean that for any state in $\mathcal{H}_{\mathsf{code}}$, the following two operations produce the same state: (a) acting with $\mathcal{O}_{\mathsf{code}}$ and then encoding with V and (b) encoding with V first and then acting with $\mathcal{O}_A$.

To see that (3.12) works, we note that reversiblity of the channel implies that there is an isomorphism $\mathcal{H}_A \cong \mathcal{H}_1 \otimes \mathcal{H}_2 \oplus \mathcal{H}_3$ where $\mathcal{H}_1 \cong \mathcal{H}_{\mathsf{code}}$ such that any state $|\psi\rangle \in \mathcal{H}_{\mathsf{code}}$ is mapped by V to the state $|\psi\rangle|\chi\rangle \in \mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \mathcal{H}_{\bar{A}}$, and where $|\chi\rangle \in \mathcal{H}_2 \otimes \mathcal{H}_{\bar{A}}$ is some fixed state.[19]

We can now evaluate (3.12). The encoding map acts as

$$V \mathcal{O}_{\mathsf{code}} V^\dagger = \mathcal{O}_{\mathsf{code}} \otimes |\chi\rangle\langle\chi|. \tag{3.14}$$

---

[19] $\mathcal{H}_3$ is just there to make sure the dimensions work out. It's unimportant.

Hence

$$\mathrm{Tr}_{\bar{A}}(V\mathcal{O}_{\mathsf{code}}V^{\dagger}) = \mathcal{O}_{\mathsf{code}} \otimes \chi_2, \tag{3.15}$$

where $\chi_2 = \mathrm{Tr}_{\bar{A}}|\chi\rangle\langle\chi|$. Finally

$$\sigma_A = \mathrm{Tr}_{\bar{A}}(\mathbb{1}_1 \otimes |\chi\rangle\langle\chi|) = \mathbb{1}_1 \otimes \chi_2. \tag{3.16}$$

Combining (3.15) and (3.16), we find $\mathcal{O}_A = \mathcal{O}_{\mathsf{code}} \otimes \mathbb{1}_2$, as desired.

For channels that are not perfectly reversible, the Petz map will not work perfectly. However, it still achieves very good performance, with an "average error" at most twice the optimal average error (using any recovery channel) [63]. For large code spaces, the performance may be suboptimal for particular "worst-case" input states, whereas recent work has shown that a slightly more complicated "twirled Petz map" has good provable worst-case performance [64]. However, for our application, the ordinary Petz map will work well enough.[20]

For reversible channels, the Petz map reconstruction of a unitary code space operator is itself always a unitary operator (at least when restricted to the image of the code space).[21] However, the individual elements that make up the Petz map reconstruction, in particular the factors of $\sigma_A^{-1/2}$, are highly nonunitary.[22] The Petz map reconstruction may therefore have much higher complexity, as a unitary operator, than the simple description (3.12) would suggest. In fact, for evaporating black holes, there is good reason to think that it is exponentially complicated [65, 66].

3.4 Gravitational replica trick for the Petz map

The above discussion of the Petz map motivates us to amend the naive "Petz Lite" formula to

$$\mathcal{O}_{\mathsf{R}} = \sigma_{\mathsf{R}}^{-1/2} \mathrm{Tr}_{\mathsf{B}}(\mathcal{O}) \sigma_{\mathsf{R}}^{-1/2}. \tag{3.17}$$

Here σ_{R} is the trace over B of the projector onto the code subspace,

$$\sigma_{\mathsf{R}} = \sum_{a=1}^{d_{\mathsf{code}}} \mathrm{Tr}_{\mathsf{B}}|\Psi_a\rangle\langle\Psi_a| = \sum_{i,j=1}^{k} |i\rangle\langle j|_{\mathsf{R}} \sum_{a=1}^{d_{\mathsf{code}}} \frac{\langle\psi_{aj}|\psi_{ai}\rangle_{\mathsf{B}}}{k}. \tag{3.18}$$

To check how well (3.17) works, we will compute the l.h.s. of (3.6) in gravity and compare it with the r.h.s. The main complication in this calculation is the fact that it involves the $-1/2$ power of an operator. To deal with this, we will use a version of the replica trick, defining

$$\mathcal{O}_{\mathsf{R}}^{(n)} = \sigma_{\mathsf{R}}^{n} \mathrm{Tr}_{\mathsf{B}}(\mathcal{O}) \sigma_{\mathsf{R}}^{n}. \tag{3.19}$$

For positive integer n, matrix elements of this operator can be evaluated using replicas. In a first pass, we will take a naive approach and analytically continue the dominant term in the bulk calculation from integer n to $n = -1/2$. Later, we will show how to do this analytic continuation more precisely, using a summation of planar geometries.

---

[20]In fact, it works better on average than the twirled Petz map.

[21]For approximately reversible channels, it will be approximately unitary.

[22]For the Petz Lite reconstruction, the problematic nonunitary part is the large constant factor c_0.

We will now begin the calculation. After writing out the integer n expression and evaluating the inner products in the R system explicitly, one gets an expression that depends only on inner products of the $|\psi_{ai}\rangle$ states in the B system:

$$\langle\Psi_a|\mathcal{O}_{\mathsf{R}}^{(n)}|\Psi_b\rangle = \frac{1}{k^{2n+2}}\langle\psi_{ai_0}|\psi_{bj_0}\rangle_{\mathsf{B}} \times \langle\psi_{a_1i_1}|\psi_{a_1i_0}\rangle_{\mathsf{B}} \ldots \langle\psi_{a_ni_n}|\psi_{a_ni_{n-1}}\rangle_{\mathsf{B}} \times \langle\psi_{b'j_n}|\psi_{a'i_n}\rangle_{\mathsf{B}}$$
$$\times \langle\psi_{b_nj_{n-1}}|\psi_{b_nj_n}\rangle_{\mathsf{B}} \ldots \langle\psi_{b_1j_0}|\psi_{b_1j_1}\rangle_{\mathsf{B}} \times \mathcal{O}_{a'b'}. \tag{3.20}$$

In this expression, all indices except for a and b are summed over. The pattern of index contractions is best understood by writing a diagram for the boundary conditions for the corresponding bulk path integral. Using the translation of inner products to boundary conditions shown in (3.2), and dropping the "$\times$" symbols for clarity, we find

$$\langle\Psi_a|\mathcal{O}_{\mathsf{R}}^{(n)}|\Psi_b\rangle = \frac{1}{(kZ_1)^{2n+2}} \times \quad\text{}\quad \times \mathcal{O}_{a'b'}. \tag{3.21}$$

In this diagram, and in similar ones below, we draw the case $n = 2$. In general there will be $2n + 2$ black segments corresponding to asymptotic boundaries.

In the phase $k \gg e^{S_{\mathrm{BH}}}$ where the connected geometries dominate, the boundary conditions will be filled in by a completely connected geometry


$$\tag{3.22}$$

We now need to evaluate this bulk computation. There are $2n + 2$ index loops for the k states of the EOW brane, which gives a factor of k^{2n+2} that cancels against the k^{2n+2} in the denominator of (3.21). What remains is a purely gravitational calculation, involving a path integral over gravity and over the propagating bulk fields.

As a first approximation (order G_N^{-1}), we can ignore the matter fields altogether and just work out the answer from the gravity computation. In this approximation, the geometry has a $\mathbb{Z}_{2n+2}$ symmetry and we can analytically continue the $\mathbb{Z}_{2n+2}$ quotient of the geometry in n. In the limit $n \to -1/2$, it becomes a disk with a single boundary, for which the gravitational path integral gives Z_1. This cancels against the Z_1^{2n+2} in the denominator, so the gravitational answer is simply one.

3.5 Reducing to a bulk field theory calculation

Now that we have evaluated the leading-order gravitational contribution, we need to compute the contribution from the propagating bulk fields. In the leading approximation, this means

a field theory calculation on the fixed background, with the $\mathbb{Z}_{2n+2}$ symmetry. We can view the geometry as built out of $2n + 2$ pieces, which we separate with dotted lines:

$$(3.23)$$
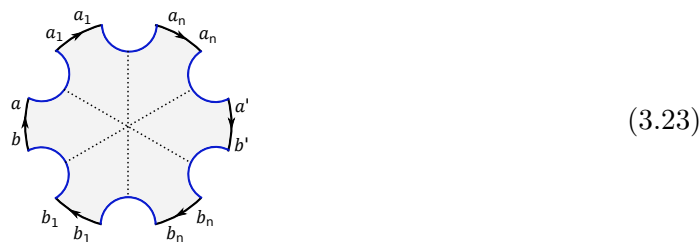
The field theory path integral on each of these pieces gives an operator with bra and ket corresponding to the dotted lines that border each piece. The particular operator we get in each case depends on the angle θ between the dotted lines, and it also depends on the field theory boundary conditions at the asymptotic boundaries, in particular, it depends on the a and b indices in (3.23). We will refer to this operator as M, and a certain sum over such operators as $\widehat{M}$:

$$M(a, b; \theta) = {}_b^a \underbrace{}_{} \theta \!\!\!\!\! , \qquad \widehat{M}(\theta) = \sum_{a=1}^{d_{\mathsf{code}}} M(a, a; \theta). \qquad (3.24)$$

In this notation, the field theory computation we are interested in is

$$\langle \Psi_a | \mathcal{O}_{\mathsf{R}}^{(n)} | \Psi_b \rangle = \sum_{a'b'=1}^{d_{\mathsf{code}}} \mathrm{Tr}\Big[M(a, b, \theta) \widehat{M}(\theta)^n M(b', a', \theta) \widehat{M}(\theta)^n \Big] \mathcal{O}_{a'b'} + O(G_N). \qquad (3.25)$$

The trace in this expression is over the Hilbert space of the bulk fields on one of the dotted lines. Since $\mathcal{O}_{\mathsf{R}}^{(n)}$ involves $2n + 2$ replicas, we should set $\theta = 2\pi/(2n + 2)$.

There are two sources of n-dependence in (3.25): the number of powers of $\widehat{M}$, and the value of θ. We would like to continue to $n = -1/2$, and it is convenient to do this in stages. In the first stage, we set $\theta = 2\pi$, leaving the number of powers of $\widehat{M}$ general. When $\theta = 2\pi$, the M operator becomes a path integral on the standard unbackreacted geometry,

$$M(a, b, 2\pi) \;=\; \raisebox{-1em}{\includegraphics{}} \;=\; \mathrm{Tr}_{\mathsf{outside}} |b\rangle \langle a|_{\mathsf{FT}}. \qquad (3.26)$$

In deriving the second equality here, we used (3.3). The notation $\mathrm{Tr}_{\mathsf{outside}}$ means a trace over the region outside the horizon, in the Hilbert space of the bulk field theory. Substituting this in, we find that with $\theta = 2\pi$, the r.h.s. of (3.25) can be written as $\langle a | \mathcal{O}_{\mathsf{inside}}^{(n)} | b \rangle_{\mathsf{FT}}$ where

$$\mathcal{O}_{\mathsf{inside}}^{(n)} = \sigma_{\mathsf{inside}}^n \mathrm{Tr}_{\mathsf{outside}} (\mathcal{O}_{\mathsf{FT}}) \sigma_{\mathsf{inside}}^n \qquad (3.27)$$

$$\sigma_{\mathsf{inside}} = \mathrm{Tr}_{\mathsf{outside}} \sum_a |a\rangle \langle a|_{\mathsf{FT}}. \qquad (3.28)$$

As a second stage, we now take the remaining dependence on n to $-1/2$, and we note that the answer takes the form of a Petz map, but now for a simpler problem defined entirely

within the Hilbert space of the bulk field theory on the fixed background. Specifically, it is the Petz map $\mathcal{O}_{\mathsf{FT}} \to \mathcal{O}_{\mathsf{inside}}$ for reconstructing the bulk field theory operator $\mathcal{O}_{\mathsf{FT}}$ using only the region inside the horizon. If $\mathcal{O}_{\mathsf{FT}}$ acts behind the horizon, then this reconstruction is possible, so this auxiliary Petz map will succeed, meaning that

$$\langle a|\mathcal{O}_{\mathsf{inside}}|b\rangle_{\mathsf{FT}} \approx \mathcal{O}_{ab}. \tag{3.29}$$

This follows from the properties of the Petz map discussed in section 3.3, but it can also be shown directly as follows. Suppose that $\mathcal{O}_{\mathsf{FT}} = \Pi_{\mathsf{code}} A_{\mathsf{inside}} \otimes \mathbb{1}_{\mathsf{outside}} \Pi_{\mathsf{code}}$, where A_{inside} commutes with Π_{code}. Then one can check that $\mathcal{O}_{\mathsf{inside}} = A_{\mathsf{inside}}$, and the reconstruction works perfectly.

Let's now summarize. We were originally interested in a reconstruction problem where the goal is to reconstruct a bulk field theory operator $\mathcal{O}_{\mathsf{FT}}$ using an operator $\mathcal{O}_{\mathsf{R}}$ acting only on the "radiation" system R. We started with a candidate for $\mathcal{O}_{\mathsf{R}}$ suggested by the Petz map. Using a bulk argument, we showed that

$$\langle\Psi_a|\mathcal{O}_{\mathsf{R}}|\Psi_b\rangle = \langle a|\mathcal{O}_{\mathsf{inside}}|b\rangle_{\mathsf{FT}} + O(G_N), \tag{3.30}$$

where the matrix element in the r.h.s. is defined purely in the bulk field theory on a fixed background. The operator $\mathcal{O}_{\mathsf{inside}}$ can be interpreted as the result of an auxiliary Petz map problem defined entirely in the field theory on the fixed background. This auxiliary Petz map attempts to reconstruct $\mathcal{O}_{\mathsf{FT}}$ using only the region behind the horizon. Such reconstruction will obviously be possible if $\mathcal{O}_{\mathsf{FT}}$ acts only behind the horizon, so (3.29) will hold and therefore (3.6) will hold also. So we conclude that in the island phase, reconstruction of operators behind the horizon is possible using only R.

3.6 The disconnected phase

To understand this result a little better, we can ask why reconstruction fails if we are in the disconnected phase with $k \ll e^{S_{\mathrm{BH}}}$. In this phase, the geometry that dominates (3.20) is the completely disconnected one, similar to the geometry shown in the center of figure 4. At leading order, the gravitational part of this computation reduces to Z_1^{2n+2}, with an additional factor of k for the single k-index loop. These factors cancel the prefactors in (3.21) in the limit $n \to -1/2$. So the gravitational and k-index part of the computation gives one, just as in the connected phase.[23] The computation therefore reduces again to a purely field-theory calculation on a fixed background.

Formally, this field theory calculation can be interpreted as an auxiliary Petz-map calculation, but of a rather trivial kind. The analog of the geometries in (3.26) is the one shown on the r.h.s. of (3.26), but with no cut at all. As a field theory operator, this corresponds to

$$\mathrm{Tr}_{\mathsf{entire\ bulk}}|b\rangle\langle a|_{\mathsf{FT}}. \tag{3.31}$$

---

[23]One could ask the following question: if in the limit $n \to -1/2$, the purely gravitational part of the answer is one for both the connected and the disconnected geometries, then should we sum both? Briefly, the answer is no. A full analysis including the transition is in section 3.7 below.

So we replace $\text{Tr}_{\text{outside}}$ by a trace over the entire bulk Hilbert space. The associated Petz map is one that attempts to reconstruct the operator $\mathcal{O}_{\text{FT}}$ from the trivial subsystem of the bulk theory, after tracing out everything. This is obviously not possible, so reconstruction fails.

3.7 Planar resummation for the Petz map

In the above analysis of the Petz map, for the most part we assumed that the connected geometry dominates. In this section, we will do the full sum over planar geometries, using a method similar to the one we used to compute the entanglement spectrum of ρ_{R}.

In this section, we will remove the bulk fields from the theory, and go back to considering the theory with only the EOW brane. In order to have a nontrivial code subspace of states

$$|\Psi_a\rangle = k^{-1/2} \sum_{j=1}^{k} |\psi_{aj}\rangle_{\text{B}} |j\rangle_{\text{R}}, \qquad a = 1, \dots, d_{\text{code}}, \tag{3.32}$$

we will interpret both the a and j indices as describing the state of the EOW brane. The j index is entangled with the radiation, and the a index labels the state within the code subspace. To reiterate: in our previous discussion, this a index labeled some state of the bulk fields, but to make the following analysis simple we will now take a to describe another aspect of the EOW brane (e.g. the state of some qubit that propagates right alongside the EOW brane).

The main challenge in the computation is the existence of the operators $\sigma^{-1/2}$ in the formula for the Petz map. It will be convenient to write a version of this formula for integer powers of σ as

$$\mathcal{O}_{\text{R}}^{n_1,n_2} = \frac{1}{d_{\text{code}}} \sigma_R^{n_1} \text{Tr}_B(\mathcal{O}) \sigma_R^{n_2}, \qquad \sigma_{\text{R}} = \frac{1}{d_{\text{code}}} \sigma_{\text{R}}. \tag{3.33}$$

In the rest of this section, we will use the rescaled σ_R defined here, which is normalized as a density matrix. What we would like to do is make sense of the "analytic continuation to $n_1 = n_2 = -1/2$" of this expression. To make this precise, it will be helpful to introduce a generating function of these operators for all values of n_1, n_2 as $\mathcal{O}_{\text{R}}(\lambda_1, \lambda_2)$:

$$\mathcal{O}_{\text{R}}(\lambda_1, \lambda_2) = \sum_{n_1,n_2=0}^{\infty} \frac{\mathcal{O}_{\text{R}}^{n_1,n_2}}{\lambda_1^{n_1} \lambda_2^{n_2}}, \qquad \mathcal{O}_{\text{R}}^{n_1,n_2} = \oint_{\infty} \frac{\mathrm{d}\lambda_1}{2\pi i} \frac{\mathrm{d}\lambda_2}{2\pi i} \lambda_1^{n_1-1} \lambda_2^{n_2-1} \mathcal{O}_{\text{R}}(\lambda_1, \lambda_2). \tag{3.34}$$

Later we will see how to deform the integration contour to get an answer that can be continued to $n_1 = n_2 = -\frac{1}{2}$, which gives the actual Petz map.

We would like to compute the sum over planar bulk geometries of the matrix elements of $\mathcal{O}_{\text{R}}(\lambda_1, \lambda_2)$. The geometries can be divided into two classes. In the first class, the asymptotic boundary corresponding to the physical black hole (with a, b indices in (3.21)) is connected to the asymptotic boundary on the other side of the circle, where the operator $\mathcal{O}_{a'b'}$ acts. In the second class, these two asymptotic boundaries are not in the same connected component of the geometry. The contribution of the first class of geometries to the matrix element is proportional to O_{ab}, since the index lines associated to the EOW branes require $a = a'$ and $b = b'$. In the second class, the contribution is proportional to $\delta_{ab} \text{Tr} \mathcal{O}$. In other words, we have the matrix elements:

$$\langle \Psi_a | \mathcal{O}_{\text{R}}(\lambda_1, \lambda_2) | \Psi_b \rangle = c_1(\lambda_1, \lambda_2) \mathcal{O}_{ab} + c_2(\lambda_1, \lambda_2) \delta_{ab} \text{Tr} \mathcal{O}. \tag{3.35}$$
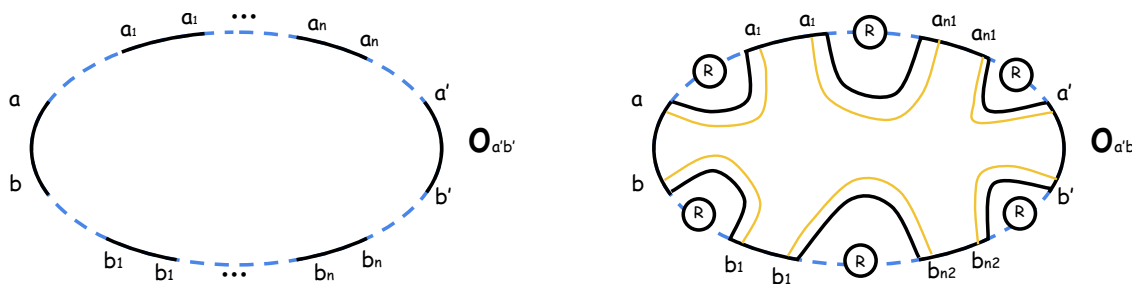
Figure 9. Boundary conditions (left) and a planar contribution to $\langle \Psi_a | \mathcal{O}_R(\lambda_1, \lambda_2) | \Psi_b \rangle$. This contribution is an element of the first class, since the leftmost boundary is in the same connected geometry as the rightmost one.

Ideal reconstruction would mean that $c_1 = 1$ and $c_2 = 0$. Since the Petz map is normalized so that the identity operator always reconstructs perfectly, we must have $c_1 + c_2 d_{\text{code}} = 1$. We can compute c_1 by summing all geometries from the first class.

Geometries belonging to the first class can be further grouped as follows. We are guaranteed that there is a geometry that connects the boundary of the physical black hole to the "opposite side" boundary. In addition, this geometry is connected to some number m_1 of boundaries in the "top half" of the circle, and m_2 boundaries in the "bottom half." This defines one connected component with $2 + m_1 + m_2$ boundaries, but in between each of these boundaries we can have a full summation of planar geometries. These can be summed over using the resolvent for σ_R:

$$\widehat{R}(\lambda) = \text{Tr} \frac{1}{\lambda - \sigma_R} \qquad \left(\text{this equals } R(\lambda) = \text{Tr} \frac{1}{\lambda - \rho_R} \text{ but with } e^{S_0} \to d_{\text{code}} e^{S_0} \right). \quad (3.36)$$

See figure 9 for a graphical description of this. The resulting expression for $c_1(\lambda_1, \lambda_2)$ in the planar approximation is

$$c_1(\lambda_1, \lambda_2) = \sum_{m_1, m_2 = 0}^{\infty} \frac{d_{\text{code}} Z_{m_1 + m_2 + 2}}{(d_{\text{code}} k Z_1)^{m_1 + m_2 + 2}} \widehat{R}(\lambda_1)^{m_1 + 1} \widehat{R}(\lambda_2)^{m_2 + 1} \lambda_1 \lambda_2. \quad (3.37)$$

In this formula, d_{code} and e^{S_0} only appear in the combination $d_{\text{code}} e^{S_0}$. This is clear for the explicit factors of d_{code}, since they multiply either Z_1 or $Z_{m_1 + m_2 + 2}$, both of which are proportional to e^{S_0}. But it is also true for the implicit factors in $\widehat{R}$, see (3.36). This will have an important implication: in order for the connected geometries to dominate the Petz map computation, we will need to have $k \gtrsim d_{\text{code}} e^{S_{\text{BH}}}$, rather than the Page time condition $k \gtrsim e^{S_{\text{BH}}}$. This is in keeping with expectations based on [13, 14, 47]. In analyzing the equations below, we will simplify by setting $d_{\text{code}} = 1$, and remembering that it can be restored by rescaling $e^{S_0} \to d_{\text{code}} e^{S_0}$.

Using the integral representation (2.32), and the notation (2.33), we can rewrite (3.37) as

$$c_1(\lambda_1, \lambda_2) = \lambda_1 \lambda_2 \int_0^{\infty} ds \, \rho(s) \frac{w(s) R(\lambda_1)}{k - w(s) R(\lambda_1)} \frac{w(s) R(\lambda_2)}{k - w(s) R(\lambda_2)}. \quad (3.38)$$

To compute the coefficient $c_1^{n_1,n_2}$, we can use the contour integral in (3.34) and interchange the s and λ integrals to write

$$c_1^{n_1,n_2} = \int_0^\infty \mathrm{d}s\,\rho(s) \oint_\infty \frac{\mathrm{d}\lambda_1}{2\pi i}\frac{\mathrm{d}\lambda_2}{2\pi i}\lambda_1^{n_1}\lambda_2^{n_2}\frac{w(s)R(\lambda_1)}{k-w(s)R(\lambda_1)}\frac{w(s)R(\lambda_2)}{k-w(s)R(\lambda_2)}. \tag{3.39}$$

For non-integer n, the function λ^n has a branchpoint at infinity, so we can't continue the integral as written. However, we can first deform the contour for the λ_1,λ_2 integrals to a new region where continuation will be possible. In particular, for each of the λ_1,λ_2 integrals, we separately deform the integration contour to surround the cut in the resolvent along the positive real axis.[24] After doing this there is no problem in continuing in n, and for $n_1 = n_2 = -1/2$, we find

$$c_1 = \int_0^\infty \mathrm{d}s\,\rho(s)\left(\oint_\mathcal{C} \frac{\mathrm{d}\lambda}{2\pi i}\frac{\lambda^{-\frac{1}{2}}w(s)R(\lambda)}{k-w(s)R(\lambda)}\right)^2. \tag{3.40}$$

3.7.1 Microcanonical ensemble

Let's first analyze this equation in the microcanonical ensemble with fixed s. Then the Schwinger-Dyson equation for the resolvent (2.34) implies

$$\lambda R = k + e^{\mathbf{S}}\frac{w(s)R}{k-w(s)R}. \tag{3.41}$$

Substituting this into the formula for c_1 (3.40), one finds

$$c_1 = e^{-\mathbf{S}}\left(\oint_\mathcal{C} \frac{\mathrm{d}\lambda}{2\pi i}\lambda^{1/2}R(\lambda)\right)^2 = e^{-\mathbf{S}}\left(\int_\mathcal{C} \mathrm{d}\lambda\,\lambda^{1/2}D(\lambda)\right)^2. \tag{3.42}$$

We can do the integral using the result for the resolvent or the density of states in (2.38), and one finds (restoring d_{code} by taking $e^{\mathbf{S}} \to e^{\mathbf{S}}d_{\mathsf{code}}$)

$$c_1 = \begin{cases} \frac{k}{d_{\mathsf{code}}e^{\mathbf{S}}} + \dots & k \ll d_{\mathsf{code}}e^{\mathbf{S}} \\ 1 - \frac{d_{\mathsf{code}}e^{\mathbf{S}}}{4k} + \dots & k \gg d_{\mathsf{code}}e^{\mathbf{S}} \end{cases}. \tag{3.43}$$

So the reconstruction works well in the region $k \gg d_{\mathsf{code}}e^{\mathbf{S}}$ where connected geometries dominate. In this region, (3.43) gives the leading small correction due to disconnected geometries.

3.7.2 Canonical ensemble

To analyze c_1 in the canonical ensemble, we can go back to (3.40) and insert the leading expression $R(\lambda) \approx k/(\lambda - \lambda_0)$ from (2.46). The λ integral is

$$\oint_\mathcal{C} \frac{\mathrm{d}\lambda}{2\pi i}\frac{\lambda^{-1/2}w(s)R(\lambda)}{k-w(s)R(\lambda)} \approx \oint_\mathcal{C} \frac{\mathrm{d}\lambda}{2\pi i}\frac{\lambda^{-1/2}w(s)}{\lambda-\lambda_0-w(s)} = -\frac{w(s)}{(\lambda_0+w(s))^{1/2}}, \tag{3.44}$$

---

[24]The denominators $k - w(s)R(\lambda)$ can be shown not to vanish on the principal sheet.
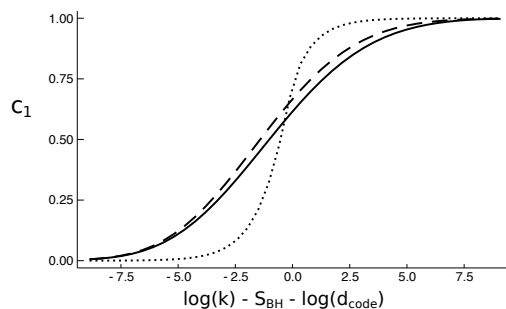
Figure 10. For $\beta = 3$ and large μ, we plot the exact c_1 in the planar approximation (solid) and the simple approximation (3.45) (dashed). The answer in the microcanonical ensemble is also shown (dotted). For smaller β, the transition in the canonical ensemble takes place over a longer interval in $\log(k)$, of order $\sim \beta^{-1/2}$.

so we find the expression

$$c_1 \approx \int_0^\infty \mathrm{d}s\, \rho(s) \frac{w(s)^2}{\lambda_0 + w(s)}. \tag{3.45}$$

Near the Page transition, the integral is dominated by the region $s < s_k$, where the $w(s)$ term in the denominator is larger than λ_0. So we can approximate the integral as

$$c_1 \approx \int_0^{s_k} \mathrm{d}s\, \rho(s) w(s) = 1 - \int_{s_k}^\infty \mathrm{d}s\, \rho(s) w(s). \tag{3.46}$$

This has a simple interpretation: part of the thermal ensemble is "pre-Page," and part of the thermal ensemble is "post-Page." The above integral gives the fraction of the ensemble that is Post-page. Note that this notion of post- and pre-Page refers to the Page time with e^{S_0} replaced by $d_{\mathsf{code}} e^{S_0}$.

After the Page transition, the correction to one in (3.46) becomes very small, and the leading correction comes from a new place: the correction to the assumed form $k/(\lambda - \lambda_0)$ of the resolvent. This correction is computed in the final term in (2.46), and working out the first-order change in the l.h.s. of (3.44), one finds

$$c_1 \approx 1 - \frac{1}{k} \int_0^\infty \!\mathrm{d}s_1 \int_0^{s_k} \!\mathrm{d}s_2\, \frac{\rho(s_1)\rho(s_2) w(s_1) w(s_2)}{(w(s_1)^{1/2} + w(s_2)^{1/2})^2}. \tag{3.47}$$

In order to simplify this answer, we set $\lambda_0 = 0$, which is a good approximation after the Page transition. This integral is dominated by the region where $s_1 \approx s_2$. Making this approximation, it then evaluates to the weighted average of the error in the microcanonical ensemble (3.43), with the weighted average taken over the Post-page portion of the thermal ensemble. This is consistent with the lower bounds on the reconstruction error derived in [13, 14].

In figure 10, we give a plot of the exact answer for c_1 in the microcanonical ensemble, and in the canonical ensemble with $\beta = 3$.

4 Radiating black holes in equilibrium

In the simple model discussed above, the Page transition results from a competition between the entropy of the black hole and the number of states k of an EOW brane. The interesting regime is a rather artificial one, where k is extremely large, of order $e^{S_{\text{BH}}}$. However, for a physical radiating black hole, we expect that a similar effect is accomplished by going to late time, where the large number of states is provided automatically by the large amount of Hawking radiation.

The analysis for a radiating black hole (even in a simple theory like JT gravity) is more complicated than for the simple model: it involves intrinsically Lorentzian physics, and the backreaction of the Hawking radiation is essential for finding the wormhole solutions. So we will not be able to go as far as for the simple model. However, in this section we will make some preliminary comments. First, we will explain the relationship between the wormhole topologies and the island extremal surface. Second, we will analyze the continuation near $n = 1$ of the replica-symmetric wormhole explicitly, and make contact with the quantum extremal surface. Finally, we will explain qualitatively why we expect solutions to exist for integer values of $n > 1$.

The setup we will discuss is the following. We consider two thermofield-double pairs of thermodynamically stable black holes (e.g. large black holes in AdS). The two pairs will be referred to as 1 and 2 — each pair consists of an L and R black hole, initially entangled in a TFD state. One can add a small interaction that couples the two L black holes togther, and likewise couples the two R black holes. A naive semiclassical computation will then show that the entanglement between system 1 and system 2 grows linearly with time forever. This eventually exceeds the coarse-grained entropy by an arbitrarily large amount. The Page curve in this context would be an entanglement entropy that follows the linear growth for a while, before saturating at a value that corresponds to the coarse-grained entropy of one pair of black holes (i.e twice the entropy of a single black hole).

4.1 JT gravity setup

We now analyze this situation in JT gravity. (2d gravity is convenient for drawing pictures, but the topological argument relating replica wormholes to the island extremal surface is similar in any spacetime dimension: one just replaces each point in the discussion below by a sphere.) We will consider two black holes in JT gravity, coupled together by a bulk matter CFT in a way that will be described below. Before getting into replicas and wormholes, let's begin by discussing the computation that gives the ordinary thermal partition function $Z(\beta)$ for this combined system.

For a single black hole, the boundary conditions (2.3) for computing $Z(\beta)$ would be to specify that the boundary is a loop of length β/ϵ and to set the dilaton to $1/\epsilon$ at this boundary. Here, $\epsilon \to 0$ is a holographic renormalization parameter. We would then look for gravity configurations that can "fill in" these boundary conditions. Up to symmetries, the unique classical solution is the hyperbolic disk

$$\mathrm{d}s^2 = \mathrm{d}\rho^2 + \sinh^2(\rho)\mathrm{d}\theta^2 = 4\frac{\mathrm{d}r^2 + r^2\mathrm{d}\theta^2}{(1-r^2)^2}, \tag{4.1}$$

where we use the portion of the geometry inside the radius

$$r = 1 - \frac{2\pi}{\beta}\epsilon. \tag{4.2}$$

This cutoff radius has been chosen so the length of the boundary is β/ϵ. The dilaton profile is

$$\phi = \phi_h \cosh(\rho) = \phi_h \frac{1 + r^2}{1 - r^2}, \qquad \phi_h = \frac{2\pi}{\beta}. \tag{4.3}$$

The value of ϕ_h was chosen so that $\phi = 1/\epsilon$ at the boundary.

Let's now discuss the computation of $Z(\beta)$ with two coupled black holes. Neglecting the matter fields for the moment, the boundary conditions will be the same in the two copies. We can fill these in with the same solution (4.1), but where we now take the two boundaries to be at radii

$$r_\pm = 1 \pm \frac{2\pi}{\beta}\epsilon. \tag{4.4}$$

The region $r < r_-$ is the first black hole, and the region $r > r_+$ is the second black hole. (Note that this is another copy of the hyperbolic disk, as one can check by taking $r \to 1/r$.) A sketch of the full configuration is here:
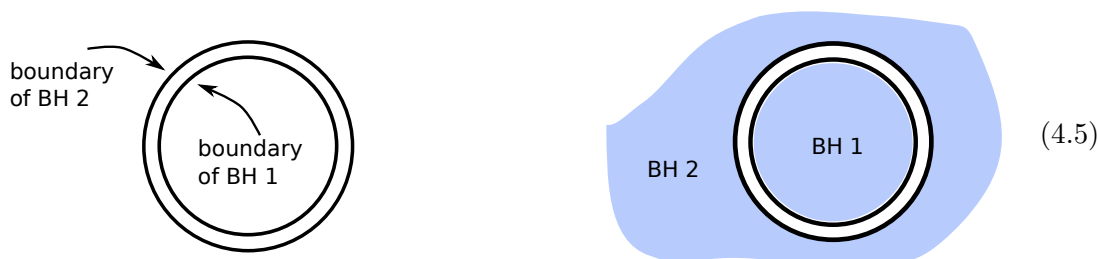

Next we should discuss the bulk matter fields and their coupling between the two black holes. A convenient choice is to include a matter CFT propagating in the bulk theory, and then allow these matter fields to pass freely from one system to the other, with "transparent" boundary conditions. So, for example, let's consider the geometry described above. For small ϵ, the geometry of the two black holes is conformal to the entire plane, and (up to a Weyl transformation that affects some observables), the bulk matter fields are simply propagating on this plane without seeing any unusual feature at $r = 1$ where we transition from one black hole to the other.

So far, we have described a situation in Euclidean signature, appropriate for computing $Z(\beta)$. But we can also continue it into Lorentzian signature. The setup corresponds to two copies of the thermofield double black hole. In other words, each of the two black holes 1 and 2 have both a left asymptotic boundary L and a right asymptotic boundary R. The boundary conditions for the matter fields allow particles to pass between the two L asymptotic boundaries, and similarly for the two R boundaries. Up to a Weyl rescaling, the bulk matter fields feel like they are propagating on a Lorentzian cylinder, where half of the spatial S^1 is in black hole 1, and the other half of the spatial S^1 is in black hole 2.

In this situation, we would like to compute the entropy of the two-sided black hole 1, namely the entropy of the subsystems $1L \cup 1R$. Let's start by considering the boundary

conditions to compute the Renyi n-entropy in the $t = 0$ state prepared by cutting the Euclidean path integral in half. These boundary conditions correspond to n replicas of (4.5), with boundary twist operators inserted, shown as zigzag cuts here:


$$(4.6)$$
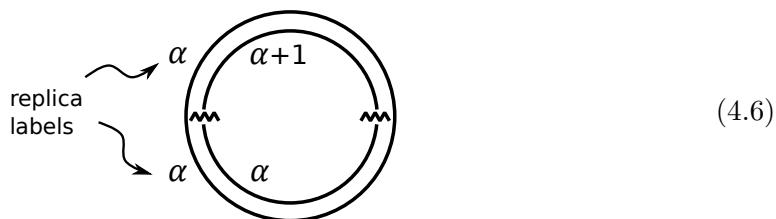
As we pass through these twist operators from below to above, we transition from replica α of the inner circle to replica $\alpha + 1$. If the systems were not interacting with each other, this would be a trivial relabeling. But with interactions it becomes nontrivial: below the twist operators replica α of the system 2 (outer circle) interacts with replica α of system 1 (inner circle), and above them replica α of system 2 interacts with replica $\alpha + 1$ of system 1.

The boundary conditions described above will compute the Renyi n-entropy at time zero of the two-sided BH 1. By moving the location of the twist operators, we can compute the entropy at other times. In principle, we can move the two twist operators independently. From the rotational symmetry of (4.6), it is clear that if we rotate the twist operators around the circle by the same angle, we will not change the answer. In Lorentzian signature, this corresponds to invariance of the answer under forwards evolution on the R system, and backwards evolution on the L system. However, if we move *both* forwards in Lorentzian time, there is nontrivial time-dependence.

4.2 Renyi topologies for the empty set and for the island

Having specified the boundary conditions, we can now ask what bulk geometries can fill them in. Our goal is to understand what topologies for the Renyi entropy computation lead to the "island" and "empty set" extremal surfaces in the von Neumann limit.

Let's first discuss the empty set. This arises from a Renyi entropy computation with the topology of $2n$ disk geometries: n for system 1, and n for system 2. Each replica of the boundary conditions is associated to a distinct replica in the bulk, and the different geometries are only connected by their boundary interactions. The geometry can be represented in a convenient way by filling in the interior and exterior of (4.6), and extending the cuts corresponding to the twist operators into the bulk, from one side of system 1 to the other:


$$(4.7)$$
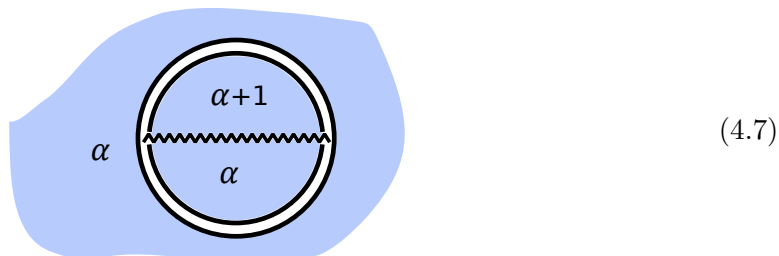
The path integral of the bulk matter fields on this geometry computes the Renyi n-entropy of the bulk matter fields on a particular bulk slice. This slice corresponds to the location of the cut in (4.7), and it can be understood as passing across the full Einstein-Rosen bridge

of system one.[25] In other words, we are computing the total entropy of the matter fields in system one. As we move the twist operators forwards in time, this entropy will grow, due to the fact that the Hawking radiation of one black hole falls into the other and vice versa. In general, the Renyi entropy computation will backreact on the gravity computation, so the geometry of the disks will be modified. But in the von Neumann limit $n \to 1$ where the entropy computation does not backreact on the geometry, we find simply

$$S_{\text{vN}} = (\text{entropy on full slice across ERB of black hole 1}). \tag{4.8}$$

In writing this expression, we used the fact that the gravitational contribution to the path integral gives one in non-backreacting limit $n \to 1$. As we move the twist operators forwards in time, the r.h.s. of (4.8) will grow linearly in time forever.

Let's now discuss the island. For the Renyi n-entropy, we leave the n disks of system 2 in place, but we replace the n disks of system 1 by a completely connected $\mathbb{Z}_n$-symmetric geometry that we will refer to as a "pinwheel." A sketch of the pinwheel for $n = 6$ (with its $\mathbb{Z}_n$ quotient shaded) looks like this:
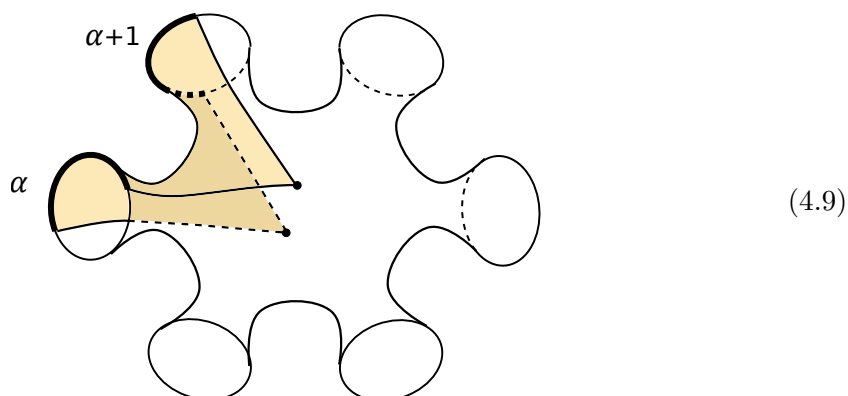
For the moment, let's not worry about whether the geometry in (4.9) is a solution or not (we will come back to this). What we would like to do is understand what it looks like in the limit $n \to 1$. Strictly speaking, this geometry does not have a continuation in n, but its $\mathbb{Z}_n$ quotient does. The quotient (shaded orange in the above sketch) is characterized by the requirement that the conical angle around the two points marked by black dots should be $2\pi/n$.

The shaded region intersects the asymptotic boundaries in two semicircles, shown with heavy black curves. These two semicircles are the two inner semicircles of (4.6), so in particular, they are joined by bulk field theory interactions to the same replica of system 2, which will be filled in by a disk geometry. For $n \approx 1$, the quotient of the full geometry including system 2 therefore looks like the following:

---
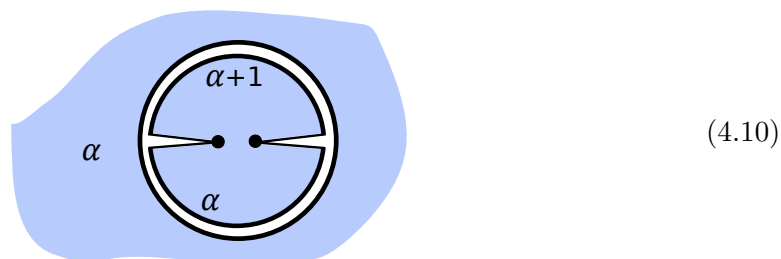
[25]This is equal to the same entropy for system two since the global state of the bulk fields is pure.

The inner region is the continuation to n near one of the shaded portion of (4.9). Note that the conical angle around the black dots is equal to $2\pi/n$, which is close to 2π.

One can evaluate the contribution of this topology to the gravity path integral, in a very crude approximation where we keep only the topological S_0 term in the JT action. In order to compute the von Neumann entropy, we need to consider

$$S_{\rm vN} = \lim_{n\to 1}\left[\frac{1}{n-1}\log\frac{\mathcal{Z}_n}{\mathcal{Z}_1^n}\right]. \tag{4.11}$$

Here $\mathcal{Z}_n$ is the bulk computation with the boundary conditions for the Renyi n entropy, and $\mathcal{Z}_1$ is the bulk computation of the partition function of the combined system. At the level of the topological action, these partition functions are simply $e^{\chi S_0}$, where χ is the total Euler characteristic:

$$\mathcal{Z}_n \sim Z_{\rm pinwheel} \times Z_{\rm disk}^n \sim e^{(2-n)S_0} \times e^{nS_0} \tag{4.12}$$
$$\mathcal{Z}_1^n \sim Z_{\rm disk}^{2n} \sim e^{2nS_0} \tag{4.13}$$

Here, the $\sim$ symbol means equivalence at the level of the topological part of the action. Substituting these expressions in (4.11), we find that the topological part of the gravitational path integral contributes $S_{\rm vN} \sim 2S_0$.

Finally, let's discuss the contribution from the bulk field theory. It is helpful to compare the island geometry in (4.10) to the empty-set one in (4.7). In the empty-set geometry (4.7), the cut associated to the twist operators on the boundary extended all the way across the bulk. In the limit $n \to 1$, the corresponding field theory computation was the entropy of the bulk fields on that slice. In the island geometry (4.10), the cut ends on the two black dots, which are the fixed points of the $\mathbb{Z}_n$ symmetry. In the limit $n \to 1$, the corresponding field theory computation is the entropy of bulk fields in the portion of system 1 outside the black dots. Adding this together with the gravity contribution described above, we have very roughly

$$S_{\rm vN} = 2S_0 + (\text{bulk entropy in BH 1 outside black dots}). \tag{4.14}$$

At the level of our topological analysis, this is consistent with the island conjecture (2.7).

In order to go beyond this crude analysis, we would like to find actual on-shell solutions in JT gravity coupled to a bulk CFT, which will require going to late times. As in many discussions of JT gravity, it is convenient to think about doing the path integral in two stages. First, we integrate over the dilaton. This imposes a delta function constraint that the metric should be hyperbolic, with $R = -2$. Next, we integrate over the moduli space of hyperbolic metrics, and over the space of "boundary wiggles," which represent the shape of the cutoff within the parent hyperbolic manifold [48]. Finding true solutions appears to involve two complications: first, in order to find on-shell solutions, we need to continue the twist operators forwards in Lorentzian time a certain amount. Second, for integer Renyi index n, the solution seems to involve an interesting but somewhat nontrivial configuration of these boundary wiggles. This problem can be avoided by continuing the off-shell action near $n = 1$, and observing that in this limit, only a finite dimensional space of wiggles can get excited. We study this problem next.

4.3 The pinwheel geometry for $n \approx 1$

Let's begin by analyzing the pinwheel geometry in more detail, and working out its off-shell action near $n = 1$. The simple final answer could be derived quickly from the methods in [7, 23], but we believe the following derivation may generalize better to integer $n > 1$.

We are interested in the special case of a pinwheel geometry with $\mathbb{Z}_n$ symmetry, and with a hyperbolic metric $R = -2$. Such geometries can be parametrized by the length b of the geodesic that separates off any one of the asymptotic regions. Other geometrical quantities can be computed in terms of this b parameter. For example, consider the two fixed points of $\mathbb{Z}_n$ symmetry, shown with black dots in (4.9). If we set 2ρ to be the geodesic distance between the two points, then one can show using hyperbolic geometry that

$$\cosh\left(\frac{b}{4}\right) = \sin\left(\frac{\pi}{n}\right)\cosh(\rho). \tag{4.15}$$

An important special case of this formula is the limit $n \to 1$, where

$$b = 2\pi \mathrm{i} - 4\pi \mathrm{i} \cosh(\rho)\,(n-1) + O((n-1)^2). \tag{4.16}$$

The pinwheel has n of these geodesics of length b, separating off the n asymptotic regions from the rest of the geometry. The asymptotic region outside each of these geodesics is a "trumpet" geometry

$$\mathrm{d}s^2 = \mathrm{d}\sigma^2 + \cosh^2(\sigma)\frac{b^2}{(2\pi)^2}\mathrm{d}\theta^2 \tag{4.17}$$

with $\sigma \geq 0$. The locus $\sigma = 0$ is the geodesic of length b, and the asymptotic boundary is at $\sigma \to \infty$.

In addition to the parameter b, the pinwheel geometry is characterized by more subtle "boundary wiggles," which represent the shape of the cutoff surface at large σ. These can be parametrized by giving the angular coordinate θ in (4.17) as a function of the proper length along the cutoff surface. It is convenient to use a rescaled proper-length coordinate τ, which runs from zero to β. The full off-shell action for the $\mathbb{Z}_n$-symmetric pinwheel geometry reduces [48] to the Schwarzian action for the trumpet geometry [25]

$$nI_{\mathrm{Sch}} = -n\int_0^\beta \mathrm{d}\tau\,\mathrm{Sch}(e^{-\frac{\theta b}{2\pi}}, \tau) = \frac{n}{2}\int_0^\beta \mathrm{d}\tau\left[\frac{\theta''^2}{\theta'^2} + \frac{b^2}{(2\pi)^2}\theta'^2\right]. \tag{4.18}$$

Putting the boundary wiggles on shell means no wiggles at all, $\theta(\tau) = \frac{2\pi}{\beta}\tau$. However, the result is still not on shell with respect to b, because one finds $I \propto b^2$.

So far, we have taken n to be general. But for n very close to one, the $\mathbb{Z}_n$ quotient of the pinwheel geometry becomes the hyperbolic disk, with two marked points corresponding to the fixed points of the $\mathbb{Z}_n$ symmetry, see (4.10). The hyperbolic disk has exact SL(2,R) symmetry, and for n close to one, we expect an approximate SL(2,R) symmetry. This leads to a separation of scales in the action: four special modes become parametrically soft. Three of these are parametrized by an SL(2,R) transformation, and one is parametrized by the distance 2ρ between the $\mathbb{Z}_n$ fixed points. We would like to work out a "slightly off-shell" action with arbitrary values of these modes, but with everything else on-shell.

A finite SL(2,R) transformation that preserves the unit circle is

$$x \to \frac{zx + y}{\bar{y}x + \bar{z}}.$$ (4.19)

The corresponding configuration of the Schwarzian variable $\theta(\tau)$ is determined by

$$e^{i\theta} = \frac{ze^{2\pi i\tau/\beta} + y}{\bar{z} + \bar{y}e^{2\pi i\tau/\beta}}.$$ (4.20)

The Schwarzian derivative $\text{Sch}(e^{-\frac{\theta b}{2\pi}}, \tau)$ for this function $\theta(\tau)$ can be computed using the composition rule $\text{Sch}(f \circ g, \tau) = \text{Sch}(f, g)g'^2 + \text{Sch}(g, \tau)$, where $f(x) = x^{\frac{ib}{2\pi}}$ and where $g(\tau) = e^{i\theta}$ given in (4.20). The integral over τ can be converted to an integral over $e^{i\theta}$, and can then be done by contour integration. The result for the action is

$$I_{\text{Sch}} = -\frac{2\pi^2}{\beta} \left[1 + \left(1 + \frac{b^2}{(2\pi)^2} \right) \frac{y\bar{y} + z\bar{z}}{y\bar{y} - z\bar{z}} \right].$$ (4.21)

Now taking the limit $n \to 1$, we can use (4.16) to expand this off-shell action near $n = 1$ as

$$I_{\text{Sch}} = -\frac{2\pi^2}{\beta} - \frac{8\pi^2}{\beta} \cosh(\rho) \frac{y\bar{y} + z\bar{z}}{y\bar{y} - z\bar{z}}(n - 1) + O((n-1)^2).$$ (4.22)

It will be convenient to think about this action in a slightly different way. We can act with an SL(2,R) transformation on the entire geometry (4.10) to "straighten out" the boundary mode (4.20). This transformation will leave the metric invariant, since the hyperbolic disk metric (4.1) is invariant under (4.19), but it will move the two marked points. A straightforward calculation shows that

$$I_{\text{Sch}} = -\frac{2\pi^2}{\beta} - 2\pi[\phi(\rho_1) + \phi(\rho_2)](n - 1) + O((n-1)^2),$$ (4.23)

where ρ_1, ρ_2 are the ρ coordinates of the new marked points, and

$$\phi = \frac{2\pi}{\beta} \cosh(\rho) = \frac{2\pi}{\beta} \frac{1 + r^2}{1 - r^2}$$ (4.24)

is the dilaton profile in the $n = 1$ geometry. This simple form of the final answer would follow more directly from the general results in [7, 23].

Let's briefly summarize. For n near one, the pinwheel geometry has four special modes with action of order $(n - 1)$. These can be parametrized by the locations of the two $\mathbb{Z}_n$ fixed points, considered as marked points in the hyperbolic disk. The off-shell action, as a function of the locations of these points, is related to the value of the dilaton at these points, (4.23).

4.4 The bulk field theory calculation

So far we have just studied the gravitational action of the pinwheel geometry. We also need to analyze the contribution from the bulk matter CFT. As explained in [8], the matter path integral can be understood as computing

$$\text{Tr}(\hat{\rho}_n^n)$$ (4.25)

where $\widehat{\rho}_n$ is the un-normalized density matrix prepared by the path integral on the $\mathbb{Z}_n$ quotient geometry. For $n \approx 1$, this approaches (4.10). In principle, we need to consider two types of corrections to $n = 1$,

$$\widehat{\rho}_n^n = \widehat{\rho}_1 + (n-1)\big[\partial_n \widehat{\rho}_n|_{n=1} + \widehat{\rho}_1 \log(\widehat{\rho}_1)\big] + O((n-1)^2). \tag{4.26}$$

The first $O(n-1)$ term can be shown to correspond to a small shift in the background value of the dilaton at the fixed points, due to quantum fluctuations of the propagating matter fields [8, 23]. This term will be subleading in the discussion below, and we will neglect it. However, the second $O(n-1)$ term leads to an important contribution for $n \approx 1$:

$$\log \mathcal{Z}_n \supset (n-1) \times (\text{bulk entropy in cut region of BH 1}). \tag{4.27}$$

We need to evaluate this entropy. This sounds potentially difficult, because the cut region in (4.10) consists of two intervals. But in the late-time limit of interest, it can be well-approximated [18] using the formula for the entropy of a single interval, which is simple in 2d CFT.

Let's review the CFT formula for the entropy of an interval. Suppose we are interested in a metric of the form

$$\mathrm{d}s^2 = \Omega(x)^2 \mathrm{d}x^i \mathrm{d}x^i, \qquad i = 1, 2 \tag{4.28}$$

We consider the state to be the vacuum in the x coordinates. For a conformal field theory, one would naively expect the entropy to be Weyl-invariant. But this is not quite true because of the need for a cutoff. If we assume that the appropriate cutoff is one in the physical distance, then the regularized entropy of an interval is (see e.g. [67])

$$S = \frac{c}{3} \log\bigg(|x - x'|\sqrt{\Omega(x)\Omega(x')}\bigg) + (\text{UV divergence ind. of } x, x'). \tag{4.29}$$

In the twist-operator formalism for the entropy, this Ω dependence comes from the Weyl-transformation properties of the twist operators.

As a warm-up, we would like to apply this to work out the regularized entropy of the interval between $r = r_1$ and $r = 1$ (at fixed θ) in the metric

$$\mathrm{d}s^2 = 4\frac{\mathrm{d}r^2 + r^2 \mathrm{d}\theta^2}{(1 - r^2)^2}, \qquad 0 \le r \le \infty. \tag{4.30}$$

This metric is conformal to the plane, and we consider the state that is the ordinary vacuum in the plane. From (4.29), the entropy of an interval $[r_1, r_2]$ at fixed θ would be

$$S_{\mathrm{bulk}} = \frac{c}{3} \log\left(\frac{r_2 - r_1}{\sqrt{(1 - r_1^2)(1 - r_2^2)}}\right) + (\text{UV divergence ind. of } r_1, r_2). \tag{4.31}$$

In the present case we are interested in taking $r_2 = 1$. As r_2 approaches 1, the expression diverges. This reflects the divergent entanglement between the two black holes due to the coupling at the boundary. This is an IR divergence in the bulk and a UV divergence in the boundary. In principle, this should be cut off by smearing out the coupling between
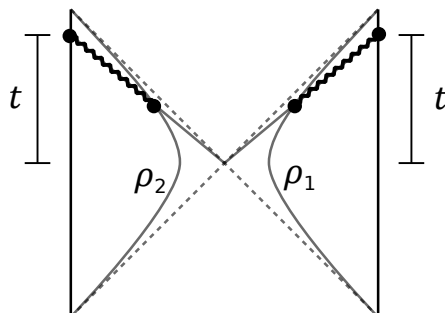
Figure 11. The black dots on the boundaries correspond to the boundary twist operators, and the black dots in the center are the marked points corresponding to the $\mathbb{Z}_n$ fixed points. For large t, the entropy on the two intervals marked with wavy lines is approximately the same as the sum of the entropies on the two intervals taken separately.

the two boundaries so that very high-energy modes are not coupled. Although we will not discuss the details of this regularization, it will lead to an expression of the form

$$S_{\text{bulk}} = \frac{c}{3} \log \frac{1 - r_1}{\sqrt{1 - r_1^2}} + \text{const} = -\frac{c}{6} \rho_1 + \text{const} \tag{4.32}$$

where the constant depends on the regularization.

Now, the entropy that we actually need to compute is the one for two intervals, where each interval stretches between one of the $\mathbb{Z}_n$ fixed points in (4.10) and one of the boundary twist operators. This problem simplifies in a late time Lorentzian configuration like the one shown in figure 11. As we increase t, the two intervals become far apart and uncorrelated with each other, and the entropy approaches the sum of two separate single-interval computations,

$$S_{\text{bulk}} \approx \left[-\frac{c}{6} \rho_1 + \text{const} \right] + \left[-\frac{c}{6} \rho_2 + \text{const} \right]. \tag{4.33}$$

In this formula, we assumed that the bulk and boundary twist operators were at the same Killing times. We won't work out the formula for different times, but we note that the action is stationary under first-order changes of these Killing times.

4.5 Saddle point

We can now put the gravity computation together with the bulk CFT computation and find a saddle point. For n near one, we have the off-shell contribution

$$\mathcal{Z}_n \approx \int \mathrm{d}^2 x_1 \mathrm{d}^2 x_2 \exp \left\{ (2 - n)S_0 - nI_{\text{Sch}} + (n - 1)S_{\text{bulk}} \right\} \tag{4.34}$$

where the integral is over the locations of the two marked points in the disk geometry. In the case where the boundary twist operators have been continued to late time, this integral has a saddle point for a Lorentzian configuration of the two marked points, corresponding to the one shown in figure 11. In principle, we should extremize the action over both coordinates. But for late time, the extremum will be such that the points x_1 and x_2 are at

the same Killing time as the corresponding boundary twist operators.[26] We then have to extremize over the radii. The action is

$$-nI_{\text{Sch}}+(n-1)S_{\text{bulk}}=n\frac{2\pi^2}{\beta}+(n-1)\left[\frac{4\pi^2}{\beta}\big(\cosh(\rho_1)+\cosh(\rho_2)\big)-\frac{c}{6}(\rho_1+\rho_2)\right]+O\big((n-1)^2\big).\tag{4.35}$$

This action has an extremum at $\sinh(\rho_1)=\sinh(\rho_2)=\beta c/24\pi^2$. This is analogous to the quantum extremal surface of [14–16, 18]. Here we have shown how it arises from a $n\to1$ limit of a Euclidean wormhole calculation.[27]

4.6 Comments on integer n

Finding saddle points explicitly for integer $n>1$ seems to be significantly more difficult than in the $n\approx1$ limit. One reason is that it is no longer possible to restrict to the SL(2,R) modes of the Schwarzian theory: a more general $\theta(\tau)$ is necessary. Gluing the two black holes together then involves a "conformal welding" problem. Without getting into the details of this, we would like to explain qualitatively why the Renyi entropy should be finite in the limit $t\to\infty$, using an argument similar to one in [24]. We caution the reader that the discussion in this section and the next one 4.7 is preliminary.

As a first step, consider the auxiliary computation

$$\text{Tr}\Big(e^{-(\beta+\beta')H_{1+2}}\Big)=\text{Tr}\Big(e^{-\beta H_{1+2}}e^{iH_{1+2}t}e^{-\beta'H_{1+2}}e^{-iH_{1+2}t}\Big)=\text{}.\tag{4.36}$$

Here H_{1+2} represents the Hamiltonian of the combined boundary dual of BH 1 and BH 2, including the interaction. In the first equality, we inserted some cancelling factors of $e^{\pm iHt}$, and in the figure at right we sketched a path integral contour that would compute this quantity. The solid and dashed lines represent the boundary duals of the two BH systems. The "caps" at the left and right end represent the Euclidean evolution by β and β', and the long horizontal portions represent the forwards and backwards Lorentzian evolution.

This quantity is exactly independent of time, and it is helpful to imagine taking t to be very large. Then the contour has a time-translation invariance that is broken only near the endpoints. The gravity solution that fills this in will also have this property, and in order for the answer to be independent of t, this time-translation-invariant gravity geometry must have zero action per unit time. In fact, the relevant geometry is easy to identify; it is a piece of a thermofield-double-like geometry, which is invariant under forwards evolution on one side and backwards evolution on the other. The action is zero per unit time because of a cancellation between the two sides.[28]

---

[26]This is because the calculation is invariant under time-reversal symmetry, once we ignore correlations between the two intervals.

[27]In the setup described here, both entangled systems are gravitational, and we could consider another saddle point where BH 1 remains disk-like and BH 2 forms a pinwheel. By making the S_0 parameters of the two theories different, we can make one of these configurations dominate over the other.

[28]We say TFD-like, because if $\beta\neq\beta'$, the two boundaries will not be on opposite sides, but at some general angle. Also, this TFD-like geometry actually consists of two TFD-like geometries, one for each of BH 1 and BH 2. However, because of the interactions between the quantum fields, it is not exactly the same as two separate TFDs; in particular, the quantum state of the matter fields will be somewhat entangled.
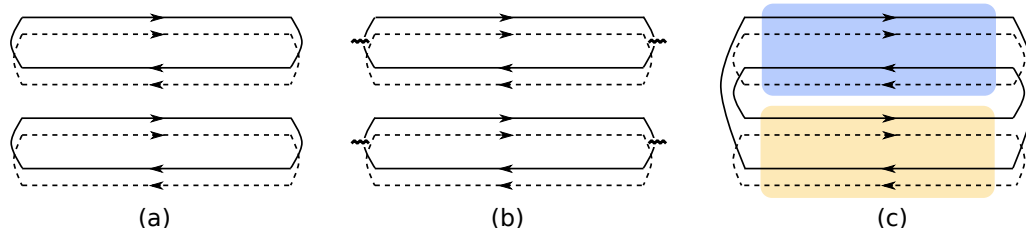
Figure 12. In (a) we sketch the boundary time contours for two copies of (4.36), where the long horizontal part is the Lorentzian portion and the "caps" at the ends are Euclidean. In (b) we insert boundary twist operators (wiggly lines) in order to compute the purity $\mathrm{Tr}(\rho_1^2)$. In (c) we represent these twist operators explicitly by changing the pattern of connection of the contours. We can make a configuration with bounded action as $t \to \infty$ by pasting the solution from the corresponding region of (a) into the shaded regions of (c). The configuration is thermofield-double-like in these two regions.

Now let's use this auxiliary computation. The contour for the Renyi 2-entropy is shown in figure 12(b) and 12(c). The long Lorentzian part of this contour is exactly the same as for two copies of (4.36), shown in figure 12(a). The difference has to do with the way the contours are connected together at the ends (and also the amount of Euclidean evolution at the ends). To find a configuration whose action does not grow with time, we can paste the solution from part (a) into the shaded region of (c), and then fill in the remainder in some way that is consistent with the new boundary conditions. Since this modification happens near the ends of the contours, it will cost an amount of action that does not depend on the length of the time interval, at least in the limit of large t. In two dimensions, the topology of the resulting gravity configuration will consist of a cylinder for BH 1 and two separate disks for BH 2.

This explains why there are gravity configurations for the Renyi entropy that have bounded action as $t \to \infty$, but it doesn't show that there are classical solutions. A potential subtlety is the following. In (4.36) there are two parameters β, β'. The action along the Lorentzian part of the contours will be zero for any values of these parameters. We expect that these are stabilized to saddle point values when we take into account the action penalty from gluing in the caps at the ends. We have not shown this reliably in gravity, but we will verify it in a related context by finding numerical solutions in the SYK model in the next section.

4.7 A factorization problem

In the context of the simple model, we argued in section 2.3 that the answers from the gravity path integral must be interpreted in terms of some implicit ensemble average. In order to make this argument, we showed that the gravity path integral predicted $\langle \psi_i | \psi_j \rangle = 0$ but $|\langle \psi_i | \psi_j \rangle|^2 \neq 0$. Can one say something similar for a more physical black hole?

In the simple model, $|\psi_i\rangle$ was the state of the black hole after projecting the radiation R onto a definite state $|i\rangle$. For a radiating black hole, an analog is as follows. We consider a case with only one black hole system, which starts out in a microcanonical version of the

TFD state, called $|E\rangle$,[29]

$$|E\rangle \propto \sum_{|E_i - E| \le \Delta} |E_i\rangle_L |E_i\rangle_R. \tag{4.37}$$

To project onto a state in which the black hole radiates a particular sequence of Hawking quanta, we could act on this initial state with a sequence of annihilation operators

$$a_{i_m}(t_m) \ldots a_{i_1}(t_1) \tag{4.38}$$

where i_j refer to a particular sequence of Hawking radiation modes, and t_j are the times at which these modes are extracted. Acting with a sequence of this type will lower the energy of the black hole, and eventually lead to evaporation. However, in order to make our arguments as sharp as possible, we would like the "evaporation" process to continue forever, so we will intersperse creation operators between the annihilation operators in such a way that the energy of the black hole remains constant.

Concretely, we will evolve forwards in time on the R boundary of the two-sided state $|E\rangle$, applying a sequence of m creation and annihilation operators to various bulk field theory modes, roughly once per unit time,

$$\mathcal{O}(i_1 \ldots i_m; t_1 \ldots t_m) = a_{i_m}^\dagger(t_k) a_{i_{m-1}}(t_{m-1}) \ldots a_{i_2}^\dagger(t_2) a_{i_1}(t_1). \tag{4.39}$$

Here, for simplicity, we are taking the bulk matter theory to be a product of free field theories. It will be convenient to constrain $\mathcal{O}$ so that it has approximately zero conserved charges in the boundary theory. So, in particular, we balance the number of creation and annihilation operators so that when we act with $\mathcal{O}$ on a state, we do not change the energy significantly.

Let's choose two different operators of the type (4.39), and call them A and B. Then we can consider states obtained by acting with these operators on the R system of the two-sided state $|E\rangle$:

$$|\psi_A\rangle = A^{(R)}|E\rangle, \qquad |\psi_B\rangle = B^{(R)}|E\rangle. \tag{4.40}$$

We expect that in JT gravity coupled to a bulk free theory, the inner product between these states will either be zero or will approach zero as we increase the number of operator insertions m,

$$\frac{\langle \psi_A | \psi_B \rangle}{\sqrt{\langle \psi_A | \psi_A \rangle \langle \psi_B | \psi_B \rangle}} \to 0. \tag{4.41}$$

We can think of this inner product as the one-point function of $A^\dagger B$ in the state $|E\rangle$. As described in section 6.2 of [26], we can get a nonzero answer for the square of this one point function, by considering a Euclidean wormhole with two boundaries. What we would like to do now is argue that the gravity answer for

$$\frac{|\langle \psi_A | \psi_B \rangle|^2}{\langle \psi_A | \psi_A \rangle \langle \psi_B | \psi_B \rangle} \tag{4.42}$$

approaches a constant as the number of insertions m goes to infinity.

---

[29]For concreteness, we assume a width Δ of order the thermal scale associated to energy E.
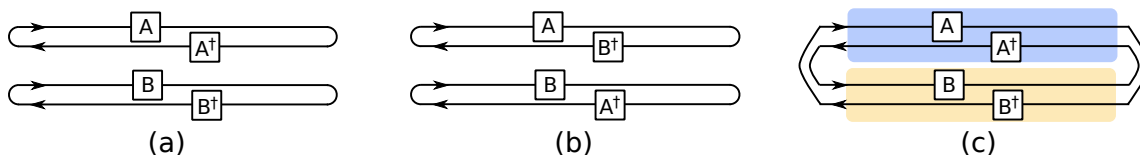
Figure 13. In (a) we sketch the boundary time contours for computing $\langle\psi_A|\psi_A\rangle\langle\psi_B|\psi_B\rangle$. In (b) we sketch the contours for computing $|\langle\psi_A|\psi_B\rangle|^2$, and in (c) we rearrange them. In the shaded regions, we glue the relevant part of the solution from (a).

The argument is very similar to the one we used in the last section to argue that Renyi entropies approach a finite answer as $t \to \infty$. The basic point is that the long Lorentzian portions of the contour (where the operators act) is the same for the numerator and denominator of (4.42), see figure 13. By gluing the gravity saddle point for the computation of the denominator into the long Lorentzian parts of the computation of the numerator, we will find a finite answer, which is suppressed only by the action involved in gluing on the different set of "caps" at the ends. The resulting topology is that of a cylinder connecting the two asymptotic boundary circles, as described in [26].[30]

5 SYK computations with two replicas

JT gravity coupled to matter fields is not a UV complete theory, due to a divergence in the moduli space integral for long thin tubes. There is no reason to suspect that this is a problem for the current calculations, but to build confidence we undertake a related calculation in the UV complete SYK model [68–71]. We will numerically find a solution that gives a non-decaying contribution to the Renyi 2-entropy, which provides a check of the argument given in section 4.6.

We study the same physical arrangement as described in section 4, but with the black hole systems 1 and 2 replaced by SYK systems 1 and 2, interacting with each other in a way that will be described below. The goal is to compute the purity $\text{Tr}(\rho^2)$, for system 1, as a function of time. Using the replica-diagonal SYK saddle point, this purity will decrease exponentially in time forever (in a microcanonical ensemble). But in the exact theory, it must saturate at an exponentially small floor value. This is a good target for a nontrivial "wormhole" saddle point, and the interchange of dominance is the Renyi 2-entropy version of the Page transition.

This SYK setup, and the replica-diagonal saddle point, was considered previously by Gu, Lucas, and Qi in [72]. In order to find the "wormhole" saddle point, we will essentially import an appropriate modification of the "double cone" solution found in [24].

To start, define two separate SYK models, called 1 and 2, with N_1 and N_2 fermions, respectively. In order to compute the Renyi 2-entropy, we will need to consider two replicas

---

[30]In the microcanonical setting described here, we expect to find a stabilized solution based on this approach. In the canonical ensemble we do not expect one, based on arguments from [24, 26].

of this system. So we will write a general fermion with three different indices, $\psi_{i,a}^{\alpha}$

$$\psi_{i,a}^{\alpha}: \quad i \in \{1, \ldots, N_a\} = \text{flavor}, \quad a \in \{1, 2\} = \text{physical system}, \quad \alpha \in \{1, 2\} = \text{replica}. \tag{5.1}$$

For each value of the replica and subsystem indices α and a, we can define an SYK Hamiltonian[31]

$$H_a^{\alpha} = \sum_{1 \leq i_1 < \cdots < i_q \leq N_a} J_{i_1 \ldots i_q; a} \ \psi_{i_1,a}^{\alpha} \ldots \psi_{i_q,a}^{\alpha}. \tag{5.2}$$

Note that the number of fermions in the two physical systems $a = 1, 2$ are in general different, and the couplings $J_{i_1 \ldots i_q; 1}$ and $J_{i_1 \ldots i_q; 2}$ are independently drawn. However, the couplings do not depend on the replica index α.

So far, the two physical systems are independent. We would like to introduce a coupling between them. In the computation of the Renyi entropy, different replicas will need to interact with each other, so we define a general operator with two replica indices α and α':

$$V^{\alpha\alpha'} = \sum_{\substack{1 \leq i_1 < \cdots < i_{\hat{q}} \leq N_1 \\ 1 \leq j_1 < \cdots < j_{\hat{q}} \leq N_2}} \hat{J}_{i_1 \ldots i_{\hat{q}}, j_1 \ldots j_{\hat{q}}} \ \psi_{i_1,1}^{\alpha} \ldots \psi_{i_{\hat{q}},1}^{\alpha} \ \psi_{j_1,2}^{\alpha'} \ldots \psi_{j_{\hat{q}},2}^{\alpha'}. \tag{5.3}$$

Regardless of the replica indices, this operator always couples the two physical systems 1 and 2 to each other, and never to themselves.[32]

Let's now specify the physical setup more precisely. The system starts out in a thermofield double state for the combined interacting $1 + 2$ system, with four subsystems $1L, 1R, 2L, 2R$. For weak interactions, this thermofield double is a state with relatively little entanglement between the 1 system and the 2 system. (Most of the entanglement is between e.g. $1L$ and $1R$, the two sides of the TFD.) We then evolve the systems forward in time, computing the entropy of system $1L \cup 1R$. The state of the combined system is invariant under forwards evolution on the R systems and backwards evolution on the L systems. We will evolve forwards on R only, leaving L alone.

After evolving for time t, we would like to compute the entropy of the $1L \cup 1R$ system. This measures the entanglement between the total 1 system and the total 2 system. We expect this entropy to grow linearly in time, before saturating at a late time value determined by a nontrivial saddle point. More precisely, at time t, we will study the purity $\text{Tr}(\rho_1(t)^2)$ of the combined $1L \cup 1R$ subsystem. This quantity can be computed by a path integral with two replicas, and with insertions of a swap operator that swaps the replica index of the $1L$ and $1R$ subsystems. We expect it to decay exponentially until saturating at a floor value.

Including the effect of the swap operators, the path integral for the purity can be written as

$$\text{Tr}\left(\rho_1(t)^2\right) = \frac{1}{Z(\beta)^2} \int \mathcal{D}\psi_{i,a}^{\alpha} e^{-I} \tag{5.4}$$

---

[31] We take q to be a multiple of four to avoid some factors below.

[32] We will take $\hat{q}$ to be even, so that V preserves separate fermion parity symmetries on the two subsystems.

where the action is [72][33]

$$I = \int_C \mathrm{d}\tau \sum_{\substack{a=1,2 \\ \alpha=1,2}} \left[\sum_{i=1}^{N_a} \psi_{i,a}^\alpha \partial_\tau \psi_{i,a}^\alpha + H_a^\alpha \right] + \int_{C_1} \mathrm{d}\tau \left[V^{11} + V^{22} \right] + \int_{C_2} \mathrm{d}\tau \left[V^{12} + V^{21} \right]. \quad (5.5)$$

Each system has two replicas, and we can either couple a particular replica of system 1 to the same replica on system 2, or to the opposite one. These two possibilities are realized on the two components of the contour $C = C_1 \cup C_2$. Switching this coupling at the transition between the two contours is equivalent to inserting swap operators at those transition points. It will be helpful below to write the full action with a condensed notation

$$\int_C \mathrm{d}\tau \sum_{\alpha\gamma} V^{\alpha\gamma} g^{\alpha\gamma}(\tau) \quad (5.6)$$

where $g(\tau)$ is the identity matrix for τ in C_1, and $g(\tau)$ is the swap operator for τ in C_2.

We can now compute the disorder average (over couplings J) of (5.4), by taking the couplings to be Gaussian random variables with mean zero and with

$$\langle J_{i_1 \ldots i_q;a} J_{i_1' \ldots i_q';a'} \rangle = J^2 \delta_{i_1 i_1'} \ldots \delta_{i_q i_q'} \delta_{aa'} \frac{(q-1)!}{N_a^q} \quad (5.7)$$

$$\langle \hat{J}_{i_1 \ldots i_{\hat{q}};j_1 \ldots j_{\hat{q}}} \hat{J}_{i_1' \ldots i_{\hat{q}}';j_1' \ldots j_{\hat{q}}'} \rangle = \hat{J}^2 \delta_{i_1 i_1'} \ldots \delta_{i_q i_q'} \delta_{j_1 j_1'} \ldots \delta_{j_{\hat{q}} j_{\hat{q}}'} \frac{(\hat{q}!)^2}{q(N_1 N_2)^{\hat{q}}}. \quad (5.8)$$

As usual in SYK calculations, the result for the disorder average can be written in terms of G, Σ collective fields[34]

$$\langle \mathrm{Tr}\left(\rho_1(t)^2\right) \rangle = \frac{1}{Z(\beta)^2} \int \mathcal{D}\Sigma_a^{\alpha\alpha'} \mathcal{D} G_a^{\alpha\alpha'} e^{-I}. \quad (5.9)$$

Note that one only needs collective fields that are diagonal in the physical system index a, although in general we need off-diagonal fields in the replica indices α, α'. The action is explicitly

$$I = \sum_a N_a \left\{ -\log \mathrm{Pf}\left(\partial_\tau \delta^{\alpha\alpha'} - \Sigma_a^{\alpha\alpha'}\right) + \frac{1}{2} \int_C \mathrm{d}\tau_1 \mathrm{d}\tau_2 \sum_{\alpha,\alpha'} \left[\Sigma_a^{\alpha\alpha'}(\tau_1, \tau_2) G_a^{\alpha\alpha'}(\tau_1, \tau_2) - \frac{J^2}{q} G_a^{\alpha\alpha'}(\tau_1, \tau_2)^q \right] \right\}$$
$$- \sqrt{N_1 N_2} \frac{\hat{J}^2}{2q} \int_C \mathrm{d}\tau_1 \mathrm{d}\tau_2 \sum_{\alpha\alpha'\gamma\gamma'} G_1^{\alpha\alpha'}(\tau_1, \tau_2)^{\hat{q}} \, g^{\alpha\gamma}(\tau_1) g^{\alpha'\gamma'}(\tau_2) \, G_2^{\gamma\gamma'}(\tau_1, \tau_2)^{\hat{q}}. \quad (5.10)$$

The saddle point equations that stationarize this action are

$$G_a = (\partial_\tau - \Sigma_a)^{-1} \quad (5.11)$$

$$\Sigma_a^{\alpha\alpha'}(\tau_1, \tau_2) = J^2 G_a^{\alpha\alpha'}(\tau_1, \tau_2)^{q-1} + \sqrt{\frac{N_{\hat{a}}}{N_a}} \hat{J}^2 G_a^{\alpha\alpha'}(\tau_1, \tau_2)^{\hat{q}-1} \sum_{\gamma\gamma'} g^{\alpha\gamma}(\tau_1) g^{\alpha'\gamma'}(\tau_2) G_{\hat{a}}^{\gamma\gamma'}(\tau_1, \tau_2)^{\hat{q}}.$$

---

[33]In reading this equation, remember that the interaction terms V always couple system 1 to system 2: the superscripts refer to the replica indices.

[34]Here we are intentionally making a small mistake and neglecting the fluctuations in $Z(\beta)$ in the ensemble of couplings. Such fluctuations are small $\sim N^{-q/2}$ and not significant for our analysis.
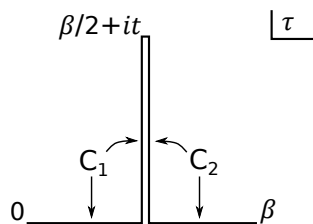
Figure 14. The time contour C, divided into C_1 and C_2. Note that $\tau = \beta$ is identified with $\tau = 0$, so C is closed. One swap operator is inserted at $\tau = 0$ and another is inserted at $\tau = \beta/2 + it$.

In the second line, we are using the notation $\hat{a}$ to mean the opposite physical system, so if $a = 1$ then $\hat{a} = 2$. As usual in SYK equations, the first equation is more complicated than it seems: for fixed a, the quantities G_a and Σ_a are viewed as matrices, acting on the vector space parametrized by τ, α. The ∂_τ operator is also viewed in these terms, but is taken to be diagonal.

These equations can be discretized on the contour 14, and solved numerically, using the standard iterative approach that is common in the SYK literature. There are two different types of solutions that will be important for us, and we will discuss them in turn.

5.1 The replica-diagonal solution

We start with the replica-diagonal solution shown in the top row of figure 15. This is the solution that was considered in [72]. As shown there, it has an action that initially grows linearly with time, implying an exponential decrease of the purity.

The growth in the action can be understood as follows. First consider the case with no interaction between system 1 and system 2, so $\widehat{J} = 0$. In this case, the twist operators have no effect, and action evaluated on the saddle point G, Σ configuration is exactly independent of time. This saddle point is simply the standard thermal solution $G(\tau)$ of the SYK model for each of the four copies (two noninteracting physical systems, with two replicas each), analytically continued along the contour C:

$$G_1^{\alpha,\alpha'}(\tau,\tau') = G_2^{\alpha,\alpha'}(\tau,\tau') = \delta^{\alpha,\alpha'} G(\tau - \tau'). \tag{5.12}$$

On this solution, if we evaluate the action with $\widehat{J} = 0$, the answer must be exactly independent of t.

Now, let's start with this solution and treat the $\hat{J}^2$ term on the second line of (5.10) as a perturbation. Evaluating the action by plugging in the unperturbed solution (5.12), the second line of (5.10) works out to

$$I \supset -\sqrt{N_1 N_2} \frac{\hat{J}^2}{q} \left[\int_{C_1} \mathrm{d}\tau_1 \mathrm{d}\tau_2 \, G(\tau_1 - \tau_2)^{2\hat{q}} + \int_{C_2} \mathrm{d}\tau_1 \mathrm{d}\tau_2 \, G(\tau_1 - \tau_2)^{2\hat{q}} \right] \tag{5.13}$$

$$= \left[\sqrt{N_1 N_2} \frac{2\hat{J}^2}{q} \int_{-\infty}^{\infty} \mathrm{d}t' G(it')^{2\hat{q}} \right] \times t + O(1). \tag{5.14}$$

Here, the $O(1)$ term represents a non-growing contribution as t becomes large. So the action will grow linearly in time, at least for while. This corresponds to an exponential decrease in the purity.
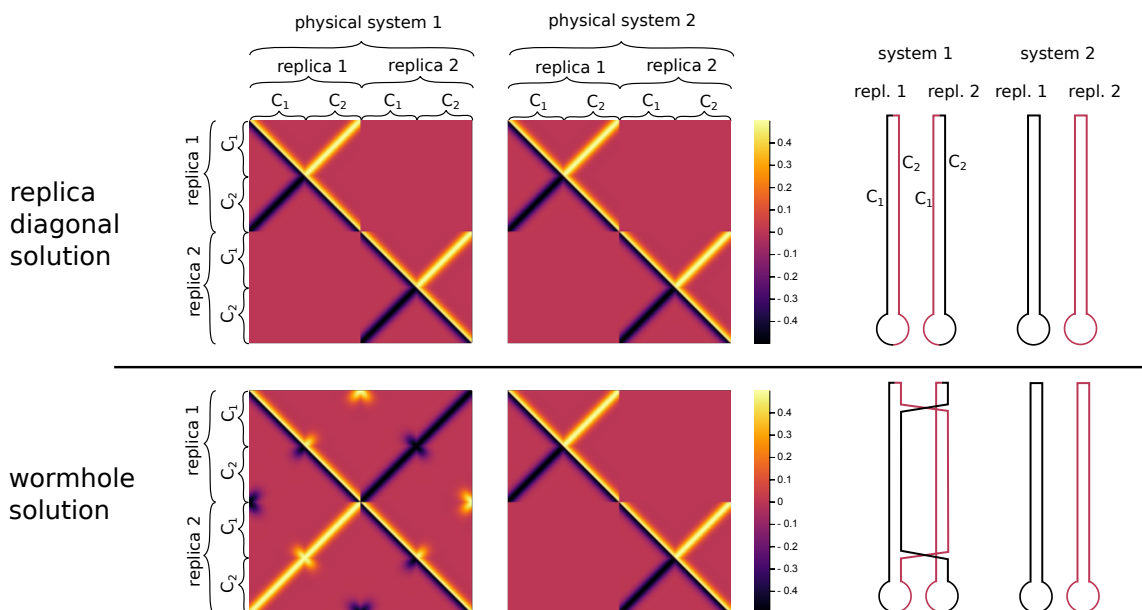
Figure 15. Numerical solutions to (5.11). The heatmaps at left represent $G_1(t, t')$ and $G_2(t, t')$ for two different types of solution discussed in the main text. We plot at infinite temperature so the solutions are real. The stripe along the main diagonal is somewhat trivial, representing large correlation between adjacent points along the contour. The other features in the heatmap represent nontrivial correlations. Contours of system 1 are interacting with the contours of system 2 with the same color due to the swap operator. The contours are drawn so that nearby points have large correlation in the solution. For this plot, $Jt = 20, \beta J = 0, \hat{J} = J/2, q = 4, \hat{q} = 2$.

This derivation is only valid for early times $\hat{J}^2 t/J \ll 1$, so that the interaction can be treated perturbatively. A subtle detail [72] is that in the canonical ensemble, this linear growth does not continue forever; instead, the action saturates at some finite value (see the left panel of figure 16). One might be tempted to interpret this saturation as the Page transition. However, what this saturation actually represents is the contribution of very low energies in the tail of the canonical ensemble, where the dynamics are slow enough that the two physical systems essentially do not evolve and entangle.[35]

This is a real effect in the Renyi 2-entropy, but such "tail" effects cannot significantly effect the von Neumann entropy (which isn't affected by tail effects the way Renyi entropies are), so we view it as a distraction. In order to avoid this subtlety altogether, we can work in the microcanonical ensemble. Then one finds that the action continues to increase linearly, as shown in the right panel of 16. So we conclude that in the microcanonical ensemble, the replica-diagonal saddle point gives an answer for the purity that exponentially decays forever.

5.2 The wormhole solution

The other class of solutions can be motivated by the discussion in section 4.6, which suggests that there should be a solution with an approximately time-translation invariant portion

---

[35]At the quantum mechanical level, the contribution of such states will still decrease with time (a power-law decrease in the purity), but we do not see this at the level of the classical action.
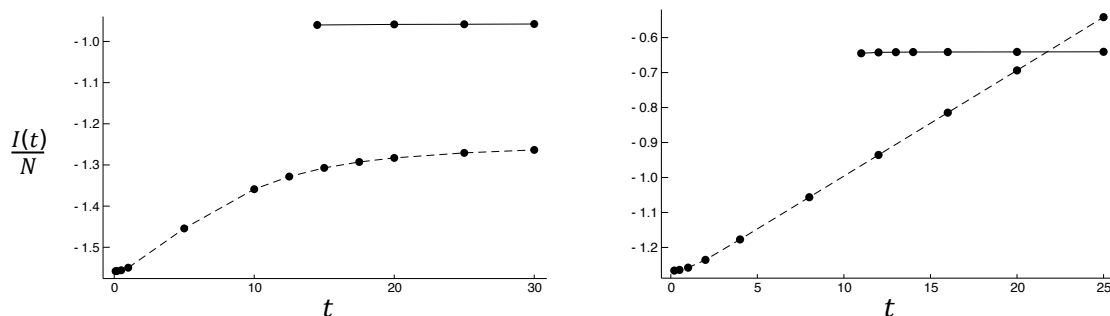
Figure 16. The action of the replica-diagonal solution (dashed) and the wormhole solution (solid) as a function of t. The left plot is in the canonical ensemble at $\beta J = 2.4$. The right plot is in the microcanonical ensemble at the corresponding energy, $E/(NJ) = -0.03$ per system. The black dots are data, and the curves are to guide the eye. The parameters were $J = 1$, $\hat{J} = J/2$, $q = 4$ and $\hat{q} = 2$.

corresponding to a TFD-like correlation between different replicas, glued in some way to a configuration with the correct boundary conditions at the ends of the contour. We refer to this as a wormhole solution because its pattern of correlation is the same as that of the wormhole geometry from section 4.6.

In order to find such a solution numerically, the procedure we followed was as follows. For the first few iterations of the Schwinger-Dyson equations, we included an explicit source in the SYK equations that encourages replica-off-diagonal correlations like the ones expected based on section 4.6. After a few iterations, we then set the source to zero and continued iterating until convergence. If the time t is larger than a critical value (discussed below), we found that the iterations converged to a nontrivial solution like the one shown in figure 15. The pattern of correlations in this solution is precisely the one expected based on the argument in section 4.6.

The numerical value of the action is independent of time to a good approximation, see figure 16.[36] To reiterate the discussion from section 4.6, this can be understood as follows: as we make the time larger, the only aspect of the solution that changes is the TFD-like portion in the middle of the time contours gets extended. Since this TFD-like configuration has exactly zero action, the action does not change.[37]

Since the action of the replica-diagonal solution increases linearly (in the microcanonical ensemble), there will eventually be a transition between the two solutions. For the setup that we have described here, the transition is at a time that is independent of N, but proportional to $1/\hat{J}^2$.

It would be desirable to understand this solution better, since it seems to involve several interesting aspects of thermalization and chaos. One example of this has to do with the critical time at which the solution first starts to exist. Empirically, this is based on the

---

[36]To compute the action accurately, we used extrapolation in the size of the matrices that represent the discretized G, Σ variables. The maximum size we used was 1600×1600 each for system 1 and system 2. To speed up convergence, one can start with a refinement of the converged solution for smaller grid size.

[37]Note that the SYK action is not local in time, but for solutions that decay exponentially in time like these, it is local enough for this argument to apply.

following: near the endpoints of the contour, the solution needs to have a "normal" pattern of correlation, in which contours C_1 and C_2 of each replica are highly correlated with each other. However, as we move away from the endpoints, this pattern of correlation is replaced by correlation between C_1 and C_2 of opposite replicas. The decay of the first type of correlation appears to be due to a scrambling effect, sourced by the perturbation due to the coupling between the systems, $\hat{J}$. Based on this, one would predict that the critical time at which the "wormhole" solutions start to exist is logarithmic in the coupling $\hat{J}$

$$t_{\text{first exist}} \propto \log(J/\hat{J}). \tag{5.15}$$

This appears to be consistent with numerics (not shown).

6 Entropy and replica wormholes in de Sitter

In this section, we will tentatively discuss some entropy computations using replica wormholes in de Sitter space.[38] Our starting point is the no-boundary proposal [73] for the wave function of de Sitter space, or more precisely its generalization in [48, 74–76] to a no-boundary proposal for the density matrix. In this version, we compute a density matrix for the universe by summing over all geometries that end on the boundary conditions for the bra and ket vectors of the density matrix. In this sum, one can have separate disconnected geometries attached to the bra and ket (these terms would also be included in the original no-boundary proposal) but also connected geometries in which the bra and ket are distinct boundaries of a single connected spacetime.

One can also generalize this further to a no-boundary proposal for the tensor product of n copies of the density matrix ρ. In this case, we have n bra-type boundaries and n ket-type boundaries, and we sum over all spacetimes (connected or otherwise) that end on these $2n$ boundary conditions.

This immediately leads to a somewhat surprising conclusion. Naively, it would appear that connected geometries will lead to a mixed density matrix. However, to check this, let's compare $\text{Tr}(\rho^n)$ and $\text{Tr}(\rho)^n$. In both cases, the boundary conditions consist of $2n$ boundary components: n ket-type boundaries and n bra-type boundaries. To compute either quantity, we identify these boundaries in bra-ket pairs, and integrate over the mutual boundary conditions for each pair. $\text{Tr}(\rho^n)$ and $\text{Tr}(\rho)^n$ correspond to two different ways of pairing up the $2n$ boundaries. However, the no-boundary rules described above are invariant under arbitrary permutations of the ket-type boundaries and the bra-type boundaries. So $\text{Tr}(\rho^n) = \text{Tr}(\rho)^n$ and the state is pure.

This seems at odds with the fact that the no-boundary answer for a single copy of the density matrix will be mixed. As in the discussion of black holes in section 2.3, a possible interpretation is that the gravity path integral is describing an ensemble of theories in which the state of the universe is pure. The average of the density matrix is the mixed no-boundary density matrix. But the average of the entropy is zero. (Another possible

---

[38] We thank Jorrit Kruthoff, Juan Maldacena, Mehrdad Mirbabayi and Eva Silverstein for discussions about de Sitter space.

interpretation is that our no-boundary prescription for $\rho^{\otimes n}$ is too aggressive. But we will keep it for the moment and see where else it leads.)

So far, we have only discussed the entropy of the whole universe. The entropy of subsystems is more interesting, but more difficult to compute. So we will study a simple model inspired by the end-of-the-world brane model for near-extremal black holes in section 2. This model is a two dimensional nearly de Sitter space described by Jackiw-Teitelboim gravity with positive cosmological constant [48, 77]. We take the asymptotic time slice to be a segment that starts and ends on two EOW branes with a large number of orthogonal internal states. The boundary conditions for an unnormalized ket vector look like the following

$$\langle ij|\Phi\rangle = \quad \overset{i}{\circ} \diagdown\!\!\!\!\!\!\!\diagup \overset{j}{\circ} \tag{6.1}$$

Here, we are imagining that we measure the length of the spatial interval to be some renormalized value ℓ, which is held fixed in the rest of the discussion. The density matrix ρ consists of two copies of this boundary condition, which can be filled in as follows:

$$= \delta_{ij}\delta_{i'j'}|\widetilde{Z}_1|^2 + \delta_{ii'}\delta_{jj'}\widetilde{Z}_2. \tag{6.2}$$

The fact that there are multiple topologies already for the density matrix makes the story in de Sitter space richer than for the black hole. Note that due to the second term, this is a mixed state. Here, in nearly de-Sitter gravity, the path integrals are given by analytic continuation of JT gravity path integrals, see [48, 77] (or more generally [78, 79]) for details of this analytic continuation. For the simple case where the brane is massless, we have

$$\widetilde{Z}_1 = 2e^{S_0}\int ds\rho(s)e^{i\ell s^2/2}|\Gamma(-\tfrac{1}{2}+is)|^2, \qquad \widetilde{Z}_2 = 4e^{S_0}\int ds\rho(s)|\Gamma(-\tfrac{1}{2}+is)|^4. \tag{6.3}$$

In order to compute e.g. $\mathrm{Tr}(\rho)$, the primed and unprimed indices will be contracted. The first diagram will contribute $e^{2S_0}k$, and the second will contribute $e^{S_0}k^2$. So if we take a very large value of k, there can be an interchange of dominance between the two.

We can now compute the entropy of the left brane degree of freedom, corresponding to the i-type index. We will refer to its density matrix as σ. For small k, a completely disconnected topology dominates, and one finds

$$\frac{\mathrm{Tr}(\sigma^n)}{\mathrm{Tr}(\sigma)^n} = \frac{k|\widetilde{Z}_1|^{2n}}{k^n|\widetilde{Z}_1|^{2n}} = \frac{1}{k^{n-1}}. \tag{6.4}$$

The von Neumann entropy will be $\log(k)$. In the opposite limit $k \gg e^{S_0}$, the fully connected geometry dominates, both for the computation of $\mathrm{Tr}(\sigma^n)$ and for $\mathrm{Tr}(\sigma)^n$. We find

$$\frac{\mathrm{Tr}(\sigma^n)}{\mathrm{Tr}(\sigma)^n} = \frac{k^{2n}\widetilde{Z}_{2n}}{k^{2n}|\widetilde{Z}_2|^n} = \frac{\widetilde{Z}_{2n}}{|\widetilde{Z}_2|^n}; \qquad \widetilde{Z}_{2n} = e^{S_0}\int ds\rho(s)(2|\Gamma(-\tfrac{1}{2}+is)|^2)^{2n}. \tag{6.5}$$

The s integral just gives an order one coefficient, so the answer is roughly $e^{(1-n)S_0}$ and the von Neumann entropy is approximately S_0.

So there is a "Page-like" transition in our simple de Sitter model. As we increase the number of EOW brane states, the von Neumann entropy of one of the branes has a transition from the naive result $\log k$ to the de Sitter entropy S_0. Although the EOW brane setup is rather artificial, this does give a hint at a microscopic role for the de Sitter entropy [80] (see also [81–84]).

It would be interesting if a similar effect could be seen in more physical setup in which the large number of states k emerges naturally from some bulk computation, rather than being put in by hand as we did here. A starting point could be the "centaur" geometry [85] or possibly the large number of states produced by quantum field theory evolution over many e-folds, which was previously considered in the context of the de Sitter entropy in [86, 87].[39]

7 Wormholes in non-averaged systems

The arguments in this paper use spacetime wormhole geometries in an essential way. But the results in section 2.3 for the overlap of individual black hole microstates $|\psi_i\rangle$ computed using such wormholes seem only to be consistent if they are interpreted as ensemble averages. In this case it seems natural to interpret the wormholes as part of an effective description, not a microscopic one. They do not know about the exponentially large amount of microscopic data contained in the fluctuating phases R_{ij} in the non-averaged matrix elements $\langle\psi_i|\psi_j\rangle = \delta_{ij} + e^{-S_0/2}R_{ij}$. We should stress that the inconsistency in overlaps that is resolved by an averaged description is not limited to the simple model discussed in section 2. As discussed in section 4.7, we expect a similar situation for the radiating black hole.

The question we want to address here is what role wormholes play in systems *without* averaging.

We would like to describe a somewhat analogous situation in semiclassical quantum chaos, which may provide some guidance [25]. Consider a few body quantum chaotic system, like a quantum billiard. Semiclassically, matrix elements in the position basis can be written as sums over classical trajectories connecting the bra point to the ket point. This should allow an analysis of the overlap puzzle. A simpler situation that illustrates many of the same ideas is to consider the Minkowski signature partition function $\mathrm{Tr}e^{-iHt}$ which can be written semiclassically as a sum over periodic orbits a. The product of two such partition functions, the spectral form factor, can be rewritten as a double sum over periodic orbits, which we can express schematically as follows:

$$K(t) = \mathrm{Tr}e^{-iHt}\,\mathrm{Tr}e^{iHt} = \sum_{ab} e^{i(S_a - S_b)}. \tag{7.1}$$

Here S_a is the classical action of orbit a.

---

[39]We than Juan Maldacena for pointing this out to us.

The spectral form factor obviously factorizes into the product of partition functions — this requires a double sum over orbits. On averaging,[40] only the "diagonal" terms corresponding to the same orbit in each sum (up to a time shift) survive. This gives the ramp.[41] After the pairing connection is made factorization is lost, as expected for an averaged system. In this picture the diagonal pairing pattern is an effective, coarse-grained description of the exponentially large number of long, diagonally paired orbits, multiplied by their exponentially small one-loop determinant.

Let's now try to make an analogy to quantum gravity. We view the quantum Hamiltonian of the billiard, H as the "boundary" description. The sum over orbits would be the microscopic "dual bulk" description. The diagonal pairing pattern in the coarse-grained sum over long orbits we take to be the analog of the wormhole geometry in the gravitational context.[42]

It is well-known that wormholes conflict with the factorization of e.g. partition functions of decoupled systems [90]. In a non-averaged situation where factorization must hold, what is one supposed to do with the wormholes? The solution the periodic orbit analogy suggests is related to one already offered in [90]. To restore factorization in the non-averaged theory, one doesn't eliminate the paired diagonal terms corresponding to the wormhole. Instead one adds back in all the other unpaired off-diagonal terms. So to have a gravitational bulk understanding of non-averaged theories we need a gravitational bulk understanding of these off-diagonal terms. These might well not have a simple geometrical description, even an effective one.

The periodic orbits are defined in the microscopic phase space that semiclassically determines all the microstates of the quantum system. So a variant of the issue at hand is to have a gravitational bulk understanding, geometrical or not, of all the microstates of the system. This is related to the fuzzball program.[43]

On occasion another idea has been suggested: wormholes in non-averaged decoupled systems could be ruled out because they are not actual saddle points. However, in JT gravity, the wormhole describing the ramp *is* a saddle point in the microcanonical ensemble [24, 26]. We see no obstruction to the existence of such microcanonically stable wormholes in more complicated higher dimensional gravitational theories dual to non-averaged systems [24]. But they would give the wrong answer in a non-averaged theory.[44] They do not describe the erratic fluctuations due to the fine-grained structure of energy levels that we expect, as illustrated for example in figure 10 of [94]. These erratic fluctuations are not a small effect — their magnitude is of the order of the signal itself. So the existence of a saddle point is not a sharp criterion for including wormholes.[45] Again, we stress that the periodic orbit analogy tells us that

---

[40]One could average over a small time intervals, for example, or over the shape of the billiard table. In the following discussion we imagine the times are long, but well before the plateau time, of order e^S.

[41]This "diagonal" approximation is due to Berry [88].

[42]Random tensor networks [89] give another example of the formation of such effective wormhole connections after averaging. The Ising domain walls discussed in [89] describe the structure of these effective connections. A closer analogy to the microstate overlaps discussed in section 2.3 would be to compute the averages of individual density matrix elements, not purities.

[43]For reviews see [91, 92]. For a critique of this program see [93].

[44]We thank Phil Saad for discussions on this point.

[45]It would be interesting to find some internal signature of this failure within the geometric bulk theory. In the presence of bulk matter there is a UV "Hagedorn" type divergence at small wormhole diameter [25] that may have some bearing on this issue.

the off-diagonal contributions are responsible for the erratic behavior. Any complete gravitational bulk description *must* contain a description of the analog of these off-diagonal terms.

In the periodic orbit example we proposed that the wormhole geometry is analogous to the diagonal pairing pattern, an effective, coarse-grained description of which orbits contribute. This raises the question of whether bulk geometry in general is only an effective description of some different, more fundamental degrees of freedom — the analog of the periodic orbits. Another possibility is that the fundamental bulk description contains "geometric" degrees of freedom, like perturbative strings, in addition to other non-geometric ones — complicated configurations of large numbers of branes, for example. It is even conceivable that geometry could actually describe everything, in some subtle way as yet not understood.[46]

The current situation, though, is that there is *no* known bulk description of a gravitational theory that is rich enough to include the microscopic information necessary to explain, for example, the erratic behavior in the spectral form factor. The nature, or perhaps even the existence, of such a bulk description remains one of the deepest mysteries in quantum gravity.

We now turn to a more pragmatic question: if spacetime wormholes are only an effective description, are such configurations useful in non-averaged bulk theories? We believe the answer is clearly "yes". For example, the entropies computed in section 2 are "self-averaging". This means that they have small variance in an averaged theory, basically because they are sums of large numbers of fluctuating terms. This variance can be computed from the bulk by considering additional wormholes linking the two copies used to compute the variance. Roughly speaking, self-averaging quantities are those where the "off-diagonal" terms make a small contribution compared to the diagonal ones.[47] We expect that a wormhole calculation of these self-averaging quantities will be quite accurate even in a single realization taken from the ensemble of theories, and hence in a non-averaged system. But we emphasize that wormholes will not give the exact answer. Worse, without understanding the bulk origin of the off-diagonal microscopic effects, there is no clear procedure to systematically improve the calculation into an arbitrarily accurate one.

Systems with a direct coupling between subsystems like those discussed in [96] are another example where self-averaging behavior occurs. Here again a wormhole calculation will be useful, even in a non-averaged theory. But again, a precise calculation would require the microscopic information.

In the absence of a complete microscopic description how is one to decide whether to trust a wormhole calculation in a non-averaged theory? An empirical test might be the following. Pretend the theory is part of an ensemble and compute the variance of the quantity of interest, again using wormholes. If it is small, the wormhole calculation should be trustworthy. If it is of order of the signal, beware!

---

[46]An impressive example of microstate information being completely described by geometry is the elegant calculation of [95] of the exact entropy of certain supersymmetric black holes. The exact integer degeneracies are recovered from a sum over an infinite number of instanton corrections computed using supersymmetric localization. What would be required in the present context is an analogous result that would explain the much more complicated pattern of energy levels present in these non BPS chaotic systems as a sum over gravitational configurations.

[47]In the periodic orbit situation, the spectral form factor averaged over a long time interval is an example of a self-averaged quantity in a non-disordered situation. Clearly the off-diagonal terms are suppressed here.

Acknowledgments

A Details on the computation of Z_n

In this appendix, we will discuss details about the calculation of I_n and Z_n. Two important formulas will be:

$$8 \int_{-\infty}^{\infty} d\ell \, K_{2is}\Big(4e^{-\frac{\ell}{2}}\Big) K_{2is'}\Big(4e^{-\frac{\ell}{2}}\Big) = \frac{\delta(s-s')}{\rho(s)}, \quad \int_{-\infty}^{\infty} d\ell \, e^{(\frac{1}{2}-\mu)\ell} K_{2is}\Big(4e^{-\frac{\ell}{2}}\Big) = \frac{|\Gamma(\mu-\frac{1}{2}+is)|^2}{2^{2\mu}}.$$
(A.1)

Using the boundary particle formalism [58, 59], the integral measure in general is the following (x_i, z_i is the location of the insertion of brane):

$$\frac{1}{2} \int_{x_1 < x_2 \ldots x_n} \frac{dx_1 dx_2 \ldots dx_n dz_1 dz_2 \ldots dz_n}{z_1^2 z_2^2 \ldots z_n^2 \mathrm{V}(\mathrm{SL}(2,\mathrm{R}))}$$
(A.2)

where z is rescaled such that the boundary is located near $z = 1$. The measure of the geodesic lengths can be determined from the above measure by first gauge fixing the SL(2,R) symmetry [58] and then a change of variable. For computation of Z_1, we need to consider $n = 2$. By gauge fixing $z_1 = z_2 = 1, x_1 = 0$, we get ($x_2 = e^{\frac{\ell}{2}}$):

$$Z_1 = \int_0^{\infty} dx_2 \, x_2^{1-2\mu} \varphi_\beta(2\log x_2) = 2e^{S_0} \int_0^{\infty} ds \rho(s) e^{-\frac{\beta s^2}{2}} \int_{-\infty}^{\infty} d\ell \, e^{(\frac{1}{2}-\mu)\ell} K_{2is}\Big(4e^{\frac{-\ell}{2}}\Big)$$

$$= e^{S_0} \int_0^{\infty} ds \rho(s) e^{-\beta \frac{s^2}{2}} 2^{1-2\mu} |\Gamma\Big(\mu - \frac{1}{2} + is\Big)|^2.$$
(A.3)

For computation of $Z_{n>1}$, we can first gauge fix $x_1 = 0, x_2 = 1, x_3 = 2$, the Faddeev-Popov determinant gives:

$$\int_{2 < x_4 < \ldots x_n} \frac{dz_1 dz_2 dz_3 dx_4 dz_4 \ldots dx_n dz_n}{z_1^2 z_2^2 z_3^2 z_4^2 \ldots z_n^2}$$
(A.4)

By defining the regularized length $\ell_{ij} = 2\log(|x_i - x_j|) - \log z_i - \log z_j$, we can introduce a new basis $\{\ell_{12}, \ell_{13}, \ell_{23} \ldots \ell_{1n}, \ell_{n-1,n}\}$, which forms a triangulation of a hyperbolic polygon:

(A.5)
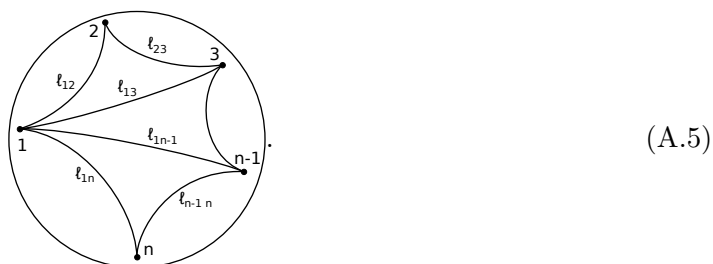
Its Jacobian matrix can be shown to be a lower triangular matrix:

$$\mathbf{J} = \left[\begin{array}{c|c} A & 0 \\ \hline \# & S \end{array}\right]; \quad A = \begin{bmatrix} -\frac{1}{z_1} & -\frac{1}{z_2} & 0 \\ -\frac{1}{z_1} & 0 & -\frac{1}{z_3} \\ 0 & -\frac{1}{z_2} & -\frac{1}{z_3} \end{bmatrix}; \quad S = \begin{bmatrix} \frac{2}{x_{41}} & -\frac{1}{z_4} & 0 & 0 & \cdots \\ \frac{2}{x_{43}} & -\frac{1}{z_4} & 0 & 0 & \cdots \\ 0 & 0 & \frac{2}{x_{51}} & -\frac{1}{z_5} & \cdots \\ 0 & 0 & \frac{2}{x_{54}} & -\frac{1}{z_5} & \cdots \\ 0 & 0 & 0 & 0 & \cdots \end{bmatrix}, \tag{A.6}$$

whose determinant is:

$$|\det \mathbf{J}| = |\det A \det S| = |\frac{2}{z_1 z_2 z_3} \frac{2x_{31}}{z_4 x_{43} x_{41}} \cdots \frac{2x_{n-1,1}}{z_n x_{n,n-1} x_{n1}}|. \tag{A.7}$$

This gives the measure:

$$\frac{1}{2^{n-1}} \int_{-\infty}^{\infty} d\ell_{12} d\ell_{13} d\ell_{23} \ldots e^{\ell_{12}/2 + \ell_{23}/2 + \ell_{34}/2 + \ldots + \ell_{n1}/2} \tag{A.8}$$

Notice that only the exterior boundaries of the hyperbolic polygon has nontrivial measure. To get the expression of I_n, we can glue multiple copies of I_3 along the inner boundaries. Using the formula of I_3 in [58], we get the expression of I_n:

$$I_n = (\ell_{12}, \ell_{23}, \ldots \ell_{n1}) = 2^n \int_0^{\infty} ds \rho(s) K_{2is}\left(4e^{\frac{-\ell_{12}}{2}}\right) \ldots K_{2is}\left(4e^{\frac{-\ell_{n1}}{2}}\right), \tag{A.9}$$

with integral measure (2.32):

$$\int_{-\infty}^{\infty} d\ell_{12} d\ell_{23} \ldots d\ell_{n1} e^{\ell_{12}/2 + \ell_{23}/2 + \ell_{34}/2 + \ldots + \ell_{n1}/2}. \tag{A.10}$$

B General two dimensional dilaton gravity

Near-extremal black holes are universally described by Jackiw-Teitelboim gravity with negative cosmological constant, and therefore they also have replica wormhole configurations where the area of the transverse direction is the dilaton field (together with S_0). It is interesting to ask whether replica wormholes exist for more general classes of black holes, especially those far from extremality.

Assuming spherical symmetry, general gravitational systems can be reduced to two dimensional dilaton gravity [97], so we are led to consider JT gravity with a more general dilaton potential in our simple model:

$$I = -\frac{S_0}{2\pi}\left[\frac{1}{2}\int_M \sqrt{g}R + \int_{\partial M} \sqrt{h}K\right] - \left[\frac{1}{2}\int_M \phi R - U(\phi) + \int_{\partial M} \sqrt{h}\phi K\right] + \mu \int_{\text{brane}} ds' \tag{B.1}$$

Again we assume the brane follows a geodesic. And again we entangle the brane states with the radiation, and consider the behavior of the entropy of the radiation when we vary k. Note that flat space black holes (Rindler space) can be described in this framework using the special case of a constant dilaton potential.

While in general understanding euclidean wormholes requires a good understanding of off-shell geometries, in the planar limit only knowledge of the path integral on the disk

topology Z_n is required. The derivation of the Schwinger-Dyson equation still applies since it only depends on the topological action and k. Moreover in the heavy brane case, the branes become local near the boundary and Z_n becomes proportional to the disk partition function of boundary length $n\beta$.[48] Since the effects of the heavy branes are local, they cancel out in the ratio of $\frac{Z_n}{Z_1^n}$ and the result reduces to just the ratio of the disk partition functions

$$\frac{Z_n}{Z_1^n} = \frac{Z(n\beta)}{Z(\beta)^n}. \tag{B.2}$$

Writing the partition function in terms of the energy density $Z(\beta) = \int dE \widetilde{\rho}(E) e^{-\beta E}$, we can again sum over n in the Schwinger-Dyson equation (2.28) and get the resolvent equation:

$$\lambda R = k + \int dE \widetilde{\rho}(E) \frac{\omega(E)R}{k - \omega(E)R}; \quad \omega(E) = \frac{e^{-\beta E}}{Z(\beta)}. \tag{B.3}$$

In general the exact form of $\widetilde{\rho}(E)$ is not known and one can only solve the equation in the classical limit. Using the classical thermodynamic relation for general dilaton gravity (appendix E.3 in [48]):

$$E = -W(\phi_h); \quad S = S_0 + 4\pi\phi_h \tag{B.4}$$

where $W(\phi) = \int^{\phi} d\phi' U(\phi)$ is called the prepotential, we can write down the general semiclassical resolvent equation:

$$\lambda R = k + \int d\phi_h e^{S_0 + 4\pi\phi_h} \frac{e^{\beta W(\phi_h)}R}{kZ(\beta) - e^{\beta W(\phi_h)}R}; \quad Z(\beta) = \int d\phi_h e^{S_0 + 4\pi\phi_h + \beta W(\phi_h)}. \tag{B.5}$$

In the microcanonical case, we will still get Page's result. In the canonical ensemble case, one needs to analyze the equation based on $W(\phi_h)$ and we expect that most of our analysis will still apply.

C   General entanglement-wedge reconstruction using Petz

In this appendix, we will indicate how the argument in sections 3.4 and 3.5 extends to the case of general entanglement-wedge reconstruction. We start with a code space of bulk field theory excitations $|a\rangle_{\mathsf{FT}}$ around some particular background, and an operator $\mathcal{O}_{\mathsf{FT}}$ acting in this space. $|\Psi_a\rangle$ is the boundary CFT state corresponding to $|a\rangle$, and $\mathcal{O}$ is the CFT operator corresponding to $\mathcal{O}_{\mathsf{FT}}$.

If A is a subregion of the boundary theory, then the Petz map gives a guess for the reconstruction of $\mathcal{O}$ on region A:

$$\mathcal{O}_A = \sigma_A^{-1/2} \mathrm{Tr}_{\overline{A}}(\mathcal{O}) \sigma_A^{-1/2}, \qquad \sigma_A = \sum_{a=1}^{d_{\mathrm{code}}} \mathrm{Tr}_{\overline{A}} |\Psi_a\rangle\langle\Psi_a|. \tag{C.1}$$

Defining a replica version as in (3.19), we have

$$\langle\Psi_a|\mathcal{O}_A^{(n)}|\Psi_b\rangle = \sum_{a'b'=1}^{d_{\mathrm{code}}} \mathrm{Tr}_A\Big[\mathbf{M}(a,b)\widehat{\mathbf{M}}^n\mathbf{M}(b',a')\widehat{\mathbf{M}}^n\Big]\mathcal{O}_{a'b'} \tag{C.2}$$

---

[48]We consider the case where there exists an asymptotic boundary.

where

$$\mathbf{M}(a,b) = \text{Tr}_{\overline{A}}|\Psi_a\rangle\langle\Psi_b|, \qquad \widehat{\mathbf{M}} = \sum_{a=1}^{d_{\text{code}}} \mathbf{M}(a,a). \tag{C.3}$$

In the bulk dual, the r.h.s. of (C.2) is a gravitational path integral with operator insertions to prepare the different state $|\Psi_a\rangle$ in the code subspace. At order G_N^{-1}, the answer doesn't depend on these operator insertions, and it reduces to the gravitational path integral for the Renyi $(2n+2)$-entropy, which we will refer to as Z_{2n+2}^{grav}.

At order G_N^0, we have a field theory computation on this fixed background, which we can write

$$\langle\Psi_a|\mathcal{O}_A^{(n)}|\Psi_b\rangle = \frac{Z_{2n+2}^{\text{grav}}}{(Z_1^{\text{grav}})^{2n+2}} \sum_{a'b'=1}^{d_{\text{code}}} \text{Tr}_{\mathcal{A}_n}\left[M(a,b;\theta)\widehat{M}(\theta)^n M(b',a';\theta)\widehat{M}(\theta)^n\right]\mathcal{O}_{a'b'} + O(G_N). \tag{C.4}$$

This formula requires some explanation. The replica-symmetric geometry $\mathcal{M}_{2n+2}$ that dominates Z_{2n+2}^{grav} has a codimension-two surface Σ that is fixed by the cyclic replica symmetry. We divide $\mathcal{M}_{2n+2}$ into $2n+2$ equal pieces by cutting along codimension-one surfaces $\mathcal{A}_n$ that connect the $2n+2$ copies of region A on the boundary to Σ. Each surface $\mathcal{A}_n$ can be understood as a backreacted Renyi version of the Cauchy slice of the entanglement wedge of A. They intersect Σ with equally spaced angles $\theta = 2\pi/(2n+2)$. The operator $M(a,b;\theta)$ is defined as the path integral on the geometry between two of these cuts, with boundary conditions that include the operator insertions for the states $|\Psi_a\rangle$ and $|\Psi_b\rangle$. In the limit where we take the total number of replicas to one, we find the simple formula

$$M(a,b;2\pi) = \text{Tr}_{\overline{\mathcal{A}}}|a\rangle\langle b|_{\mathsf{FT}}. \tag{C.5}$$

As in section 3.5, one can now take the $n \to -1/2$ limit, and find

$$\langle\Psi_a|\mathcal{O}_A|\Psi_b\rangle = \langle a|\mathcal{O}_{\mathcal{A}}|b\rangle_{\mathsf{FT}} + O(G_N) \tag{C.6}$$

where the r.h.s. is an auxiliary Petz map computation, defined purely in the bulk field theory. $\mathcal{A}$ is the Cauchy slice of the entanglement wedge of A, and

$$\mathcal{O}_{\mathcal{A}} = \sigma_{\mathcal{A}}^{-1/2}\text{Tr}_{\overline{\mathcal{A}}}(\mathcal{O}_{\mathsf{FT}})\sigma_{\mathcal{A}}^{-1/2}, \qquad \sigma_{\mathcal{A}} = \sum_{a=1}^{d_{\text{code}}} \text{Tr}_{\overline{\mathcal{A}}}|a\rangle\langle a|_{\mathsf{FT}}, \tag{C.7}$$

is the Petz map for the channel corresponding to erasure of the complement of the entanglement wedge. In particular, for the case where $\mathcal{O}_{\mathsf{FT}}$ acts within the entanglement wedge $\mathcal{A}$ to a good approximation, then $\mathcal{O}_{\mathcal{A}} = \mathcal{O}_{\mathsf{FT}}$, and reconstruction succeeds.

We also note that there is a more general version of the Petz reconstruction map, defined using a fixed, but arbitrary, state ρ that is not necessarily maximally mixed. In this case, we define

$$\mathcal{O}_A^{(\rho)} = \rho_A^{-1/2}\text{Tr}_{\overline{A}}(\rho^{1/2}\mathcal{O}\rho^{1/2})\rho_A^{-1/2}. \tag{C.8}$$

It is easy to see, by analogous arguments to those above, that this reconstruction reduces in the bulk to the field theory Petz reconstruction

$$\mathcal{O}_{\mathcal{A}}^{(\rho)} = \rho_{\mathcal{A}}^{-1/2} \text{Tr}_{\overline{\mathcal{A}}}(\rho^{1/2} \mathcal{O}_{\text{FT}} \rho^{1/2}) \rho_{\mathcal{A}}^{-1/2}. \tag{C.9}$$

An advantage of this more general construction is that it can be made well-defined even for infinite-dimensional code spaces, where the maximally mixed state does not exist. However, one has to be somewhat careful here: if the code space includes degrees of freedom outside the entanglement wedge (i.e. we have a *subsystem* rather than a *subspace* code), the field theory Petz map, constructed using an arbitrary, non-maximally mixed state ρ, will not necessarily recover the original operator. Even if we use the twirled Petz map, the reconstruction is only guaranteed to work if the code space state ρ has no entanglement between the inside and outside of the entanglement wedge [39].

In an infinite-dimensional code space (such as the entire field theory Hilbert space), such product states do not necessarily exist. However, by using a thermal state at very high temperature, we can make the entanglement arbitrarily short range. We should then expect that the field theory Petz map will recover the original operator with high accuracy so long as the inverse temperature β is much smaller than the distance from the original operator to the edge of the entanglement wedge. This can be verified in simple cases where the action of the field theory modular flow (and hence the field theory Petz map) is known explicitly.

Finally, we emphasize that the definition of the Petz map reconstructions relies on being able to create arbitrary states in the code space using gravitational path integrals. For most situations of interest, this is not a problem. For example, if our code space is the state of a diary that was thrown into a black hole, we can easily construct arbitrary states in the code space simply by changing the state of the diary, before it was thrown into the black hole.

If we want to reconstruct the interior partners of Hawking quanta — to understand, for example, whether there is a firewall [29] — the situation is somewhat more complicated. We cannot directly manipulate the interior modes because they become trans-Planckian when evolved back in time. Instead, they are always produced, together with the Hawking quanta themselves, in a fixed, entangled state $|\psi_0\rangle$. However, we can use the gravitational path integral to manipulate the state of the Hawking quanta. Because the state $|\psi_0\rangle$ has maximal rank, using these manipulations, we can produce an overcomplete basis of states for the Hawking quanta and interior partners.

By taking linear superpositions of such path integrals, we can therefore construct arbitrary states in code spaces that include interior partner modes. We can use these path integrals to construct Petz map reconstructions of interior modes, after the Page time, that act only on the Hawking radiation.

This construction assumes that the interior modes were initially in the state $|\psi_0\rangle$, i.e. that there wasn't a firewall. Indeed, it has always been the case that gravitational calculations implied the absense of a firewall. The new result that one can see using the Petz map is that gravitational calculations also imply that the interior modes can be reconstructed on the Hawking radiation, i.e. ER=EPR [30–34].

D A ensemble boundary dual of the simple model

In this appendix, we show that the simple model, with pure JT gravity plus EOW branes, is dual to a boundary ensemble of theories, just like pure JT gravity [25]. This ensemble of theories is defined by a randomly chosen Hamiltonian H, and a set of randomly chosen special states $|i\rangle$ (the brane states). The dual bulk theory is valid to all order in e^{-S_0}. It includes nontrivial bulk topologies, including topologies with handles (suppressed by powers of e^{-S_0}). However it doesn't include any contributions of EOW brane loops. All brane world lines have to begin and end on the boundary. This provides some justification for the fact that we ignored the possibility of end-of-the-world brane loops in all our bulk gravity calculations in this paper.

We now give a precise definition of the ensemble of boundary theories. First, the Hamiltonian H is chosen from the usual JT gravity ensemble of Hamiltonians, as in [25]. Let the eigenstates of this Hamiltonian be labelled $|E_a\rangle$. Then the brane states are chosen to be

$$|\psi_i(\beta)\rangle = \sum_a 2^{1/2-\mu}\Gamma\left(\mu - \frac{1}{2} + i\sqrt{2E_a}\right)e^{-\beta E_a/2}C_{i,a}|E_a\rangle \tag{D.1}$$

where the coefficients $C_{i,a}$ are i.i.d. complex Gaussian random variables.

Let us see why this works. Our aim is to show that expressions of the form

$$\mathbb{E}_{H,\{C\}}\left(\prod_{m=1}^{p}\langle\psi_{i_m}(\beta_m)|\psi_{j_m}(\beta_m)\rangle \prod_{n=1}^{q}\mathrm{Tr}(e^{-\widetilde{\beta}_n H})\right), \tag{D.2}$$

where the expectation is over the ensemble of Hamiltonians and states (D.1), matches a bulk computation in JT gravity with the following boundary conditions. We have p asymptotic boundaries that are intervals of renormalized lengths $\{\beta_m\}$ bounded by EOW branes, and we have q standard S^1 boundaries with renormalized lengths $\{\widetilde{\beta}_n\}$.

We now simply integrate out the Gaussian random variables $C_{i,a}$. We find that (D.2) equals

$$\sum_\pi\left[\prod_{m=1}^{p}\delta_{i_{\pi(m)},j_m}\mathbb{E}_H\left(\prod_{\gamma\in c(\pi)}\mathrm{Tr}\left(\prod_{m\in\gamma}\left[e^{-\beta_m H}\sum_a \frac{|\Gamma(\mu-\frac{1}{2}+i\sqrt{2E_a})|^2}{2^{2\mu-1}}|E_a\rangle\langle E_a|\right]\right)\prod_{n=1}^{q}\mathrm{Tr}(e^{-\widetilde{\beta}_n H})\right)\right] \tag{D.3}$$

Here, we are summing over permutations π on p elements that take m to $\pi(m)$ and the subsets $\gamma\subseteq\{1,2,\ldots,p\}$ are elements of the set $c(\pi)$ of cycles of the permutation π. This formula no longer involves the brane states $|i\rangle$. We can therefore hope to make contact with the arguments from [25] for evaluating products of partition functions, in this ensemble of Hamiltonians, using bulk JT gravity path integrals.

Of course, (D.3) isn't quite in the right form that would allow us to do this, since it includes insertions of $|\Gamma(\mu-\frac{1}{2}+i\sqrt{2E_a})|^2|E_a\rangle\langle E_a|$. We therefore use the inverse Laplace transform $f(\beta') = \mathrm{IL}[2^{1-2\mu}|\Gamma(\mu-\frac{1}{2}+i\sqrt{2E_a})|^2](\beta')$ to write

$$\sum_a 2^{1-2\mu}|\Gamma\left(\mu - \frac{1}{2} + i\sqrt{2E_a}\right)|^2|E_a\rangle\langle E_a| = \int d\beta' f(\beta')\sum_a e^{-\beta' E_a}|E_a\rangle\langle E_a|. \tag{D.4}$$

This turns the evaluation of (D.3) into an integral over products of partition functions, with 'inverse temperatures' that depend on the auxiliary variables β'. Explicitly, we have

$$\sum_\pi \left[\int \prod_m d\beta'_m \prod_m [f(\beta'_m)\delta_{i_{\pi(m)},j_m}] \mathbb{E}_H \left(\prod_{\gamma \in c(\pi)} \mathrm{Tr}\left(e^{-\sum_{m \in \gamma}(\beta_m + \beta'_m)H} \right) \prod_{n=1}^q \mathrm{Tr}\left(e^{-\widetilde{\beta}_n H} \right) \right) \right]. \tag{D.5}$$

The expectation values in this formula can be evaluated using bulk JT gravity by considering all topologies (including topologies with handles) of JT gravity with asymptotic boundaries of renormalised length $\sum_{m \in \gamma}(\beta_m + \beta'_m)$ for each cycle γ in the permutation π, as well as additional asymptotic boundaries of renormalised length $\widetilde{\beta}_n$ for each n.

Once we include the sum over permutations π, this is exactly the set of topologies that appear in the simple model. The different permutations correspond to the different ways of pairing up brane start and end points. The factors of $\delta_{i_{\pi(m)},j_m}$ ensure that we only get nonzero contributions when the state of each brane is the same at both ends; it tells us that the branes have no dynamics, as desired.

We are still not quite done however. In this formula, we consider geometry of fixed asymptotic length, and integrate over different lengths. In contrast, in JT gravity with branes, the topologies are partially bounded by geodesics, which contribute the EOW brane action μl. Fortunately, as argued in [26, 58], any boundary segment of length β'_m can be replaced in the JT gravity calculations by the insertion of the Wheeler-de Witt wavefunction at the homologous, non-selfintersecting bulk geodesic χ. In the l basis, this Wheeler-de Witt wavefunction is given by

$$\psi_{\beta'_m}(l) = 4e^{-l/2} \int_0^\infty ds\, \rho(s) e^{-\beta' s^2/2} K_{2is}(4e^{-l/2}), \tag{D.6}$$

as in (2.31). When we rewrite the JT gravity calculation in this way, the only dependence of the expectation value in (D.5) on the auxiliary variables β'_m comes in the choice of wavefunction that we insert at the geodesic.

We can now do the integrals over β'_m explicitly by simply taking a superposition over these Wheeler-de Witt wavefunctions. After doing this superposition, we find that the full wavefunction on each geodesic is given by

$$\psi(l) = 2^{3-2\mu} e^{-l/2} \int_0^\infty ds\, \rho(s) |\Gamma\left(\mu - \frac{1}{2} + is\right)|^2 K_{2is}\left(4e^{-l/2}\right). \tag{D.7}$$

Using (A.1) and eq. 25 of [98],

$$8 \int_0^\infty ds\, \rho(s) K_{2is}\left(4e^{-l/2}\right) K_{2is}\left(4e^{-l'/2}\right) = \delta(l - l'), \tag{D.8}$$

which is the orthogonality condition for the l-basis of wavefunctions, we find

$$\psi(l) = e^{-\mu l}. \tag{D.9}$$

This is just the insertion of a brane wavefunction on the geodesic χ. We have therefore recovered the bulk theory of JT gravity with EOW branes.

We can now use the random state formula (D.1) to write a random matrix formula for the density matrix ρ_{R} in the state (2.5):

$$\rho_{\mathsf{R}} = C F(H) C^\dagger, \qquad F(x) = |\Gamma\left(\mu - \frac{1}{2} + i\sqrt{2x}\right)|^2 e^{-\beta x}. \qquad \text{(D.10)}$$

The matrix H is drawn from the double-scaled matrix integral dual to ordinary JT gravity. The matrix C is a rectangular $k \times \infty$ complex matrix with Gaussian random entries, rescaled so that the density matrix is correctly normalized.

E   Fixed area states and random tensor networks

In this appendix, we show that the non-perturbative corrections to the Ryu-Takayanagi formula from additional extremal surfaces are identical in a) "fixed-area" states in AdS/CFT [60, 61] and b) random tensor networks [89]. This is consistent with the intuition from [60, 61] that fixed-area states are analogous to tensor network states, since both have flat entanglement spectra at leading order. However, our calculations suggest that the connection to *random* tensor networks specifically (as opposed, for example, to perfect tensor networks [99]) goes much deeper than might previously suspected, with non-perturbative instanton corrections that are the same in both models. **Note:** this appendix was added in v2; similar results were derived independently in upcoming work by Akers, Faulkner, Lin and Rath [100].

We first calculate the von Neumann entropy, including non-perturbative corrections, for the simplest possible random tensor network: a bipartite pure state with subsystem Hilbert space dimensions d_B and $d_{\bar{B}}$. (We assume for notational convenience that $d_B < d_{\bar{B}}$.) This is, of course, just the famous Page calculation [2]. Let us review how the answer can be found using the replica trick.

We first evaluate $\langle \text{Tr}(\rho_B^n) \rangle$, where the expectation is over the Haar measure for the random state. Using the standard result

$$\langle |\psi\rangle\langle\psi|^{\otimes n} \rangle \propto \sum_\pi \pi_B \pi_{\bar{B}}, \qquad \text{(E.1)}$$

where the sum is over permutations π on the n copies of the system and π_B and $\pi_{\bar{B}}$ are operators that permute the n copies of their respective subsystem, we find

$$\langle \text{Tr}(\rho_B^n) \rangle = \left\langle \frac{\text{Tr}(\rho_B^n)}{\langle\psi|\psi\rangle^n} \right\rangle = \frac{\sum_\pi \text{Tr}(\tau_B \pi_B)\text{Tr}(\pi_{\bar{B}})}{\sum_\pi \text{Tr}(\pi_B)\text{Tr}(\pi_{\bar{B}})} = \frac{\sum_\pi d_B^{C(\tau\circ\pi)} d_{\bar{B}}^{C(\pi)}}{\sum_\pi d_B^{C(\pi)} d_{\bar{B}}^{C(\pi)}}, \qquad \text{(E.2)}$$

where τ is any fixed cyclic permutation and $C(\Pi)$ is the number of cycles in the permutation Π. Here we focus on the limit where d_B and $d_{\bar{B}}$ are both large but their ratio is arbitrary. In this limit (E.2) simplifies to

$$\langle \text{Tr}(\rho_B^n) \rangle = \sum_{\pi \in G} d_B^{C(\tau\circ\pi)-n} d_{\bar{B}}^{C(\pi)-n}, \qquad \text{(E.3)}$$

where the sum is now over permutations in the set G of permutations such that $C(\pi)+C(\tau\circ\pi)$ takes its maximal value, which is $n + 1$. These are the permutations that lie on a geodesic

(shortest paths in permutation space, where each step is a transposition) between the identity permutation and the cyclic permutation τ. Such permutations are in one-to-one correspondence with the set of non-crossing partitions. We note that, in the limit of large $d_B, d_{\bar{B}}$, fluctuations are suppressed and so the expectation over states is unnecessary.

The number of non-crossing permutations such that $C(\tau \circ \pi) = k$ is equal to the Narayana number, $N(n, k)$. Substituting in the explicit expression for $N(n, k)$, we have

$$\langle \mathrm{Tr}(\rho_B^n) \rangle = \frac{1}{d_B^{n-1}} \sum_{k=1}^{n} N(n, k) \left(\frac{d_B}{d_{\bar{B}}} \right)^{k-1} = \frac{1}{d_B^{n-1}} \sum_{k=1}^{n} \frac{1}{n} \binom{n}{k} \binom{n}{k-1} \left(\frac{d_B}{d_{\bar{B}}} \right)^{k-1}. \quad \text{(E.4)}$$

To find the von Neumann entropy, we would like to analytically continue this expression in n and then expand the answer near $n = 1$. This looks awkward, because the range of the sum itself depends on n. However, we can use the fact that the binomial coefficients vanish for all $k > n$ to extend the sum from $k = 1$ to infinity. The expression is then easy to continue in n, and we find

$$\mathrm{Tr}(\rho_B^n) = \frac{1}{d_B^{n-1}} \left[1 + \frac{n(n-1)}{2} \frac{d_B}{d_{\bar{B}}} + O\left((n-1)^2 \frac{d_B^2}{d_{\bar{B}}^2} \right) \right]. \quad \text{(E.5)}$$

The first term comes from the identity permutation ($k = 1$) and the second comes from the permutations consisting of a single transposition ($k = 2$), which give the leading correction in the limit $d_{\bar{B}} \gg d_B$. We therefore find

$$S = -\mathrm{Tr}(\rho_B \log \rho_B) = -\partial_n \mathrm{Tr}(\rho_B^n)|_{n=1} = \log d_B - \frac{d_B}{2d_{\bar{B}}}. \quad \text{(E.6)}$$

Naively, one would expect additional subleading corrections from permutations that are suppressed by higher powers of $d_B/d_{\bar{B}}$, as there were for $\mathrm{Tr}(\rho_B^n)$. However, one can check that for $k > 2$, the continuation of the Narayana number $N(n, k)$ near $n = 1$ is proportional $(n-1)^2$, which does not contribute to the von Neumann entropy. Morally speaking, the von Neumann entropy is "one-instanton exact".

We note that much of the above discussion extends to more complicated random tensor networks. For example, when evaluating $\mathrm{Tr}(\rho_B^2)$, one ends up with the partition function of a classical Ising model. However, this will not be necessary for our present purposes. See [89] for more details.

Instead, we shall move on to discussing fixed-area states in quantum gravity. For concreteness, we shall mostly focus on the example of two disjoint boundary regions (which we collectively denote by B) in AdS$_3$, although the story is completely general. In this case, there are two extremal surfaces, one homotopic to the region B (with area A_1) and the other homotopic to the complementary region $\bar{B}$ (with area A_2).

We consider states where the area of both surfaces has been measured, and so saddle points of the gravitational path integral can have conical singularities at the extremal surfaces, so long as the deficit angle is constant everywhere on the surface. As discussed in [60, 61] and section 3.2, this means that the replicated geometries used to calculate $\mathrm{Tr}(\rho_A^n)$ are very simple: they are just n copies of the original unbackreacted geometry, with different sheets glued together around the extremal surfaces.
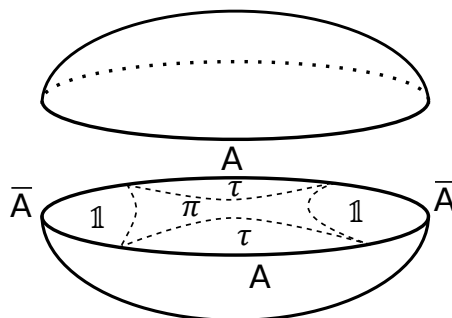
Figure 17. An example of a replicated geometry for two intervals in AdS, cut open.

More specifically, as shown in figure 17, the boundary conditions used to compute $\text{Tr}(\rho_B^n)$ mean that the bottom half (the 'ket') of the bulk geometry near region B needs to be glued to the top half of the bulk geometry on the neighbouring sheet. Passing upwards through this region applies a cyclic permutation τ to the different sheets. In contrast, when upwards through the bulk geometry near region $\bar{B}$, one must remain on the same sheet. What about the central region, between the two extremal surfaces? This is not fixed by the boundary conditions, and so there should be saddles associated to each possible permutation π of the sheets.

What is the action of each of these saddle points? We first note that, when the state is correctly normalised (dividing through by $\text{Tr}(\rho_B)^n$) the contribution to the gravitational action from everything except the conical singularities cancels between numerator and dominator because the geometry is unbackreacted. The conical singularites themselves give a contribution to the action of $(\hat{\phi} - 2\pi)A/8\pi G_N$, where $\hat{\phi}$ is the angle around the conical singularity. If the extremal surfaces had conical singularities with angles ϕ_1 and ϕ_2 in the unreplicated geometry, the contribution to the action in the replicated geometry is

$$[n\phi_1 - 2\pi C(\tau \circ \pi)]\frac{A_1}{8\pi G_N} + [n\phi_2 - 2\pi C(\pi)]\frac{A_2}{8\pi G_N}.$$

After normalisation, the dependence on ϕ_1 and ϕ_2 vanishes and we are left with the final result

$$\text{Tr}(\rho_A^n) = \sum_\pi \exp\left([C(\tau \circ \pi) - n]\frac{A_1}{4G_N} + [C(\pi) - n]\frac{A_2}{4G_N}\right). \tag{E.7}$$

Note that, since A_1 and A_2 are individually divergent, all permutations π that do not minimise $(C(\tau \circ \pi) + C(\pi))$ are infinitely suppressed, even at finite G_N.

This is exactly the result that we found for the bipartite random pure state, except with d_B and $d_{\bar{B}}$ replaced by $\exp(A_1/4G_N)$ and $\exp(A_2/4G_N)$ respectively. It therefore follows that the non-perturbative corrections to the von Neumann entropy for the fixed-area state are identical to the corrections that we found above for the bipartite random pure state.

Finally, we comment briefly on what happens when bulk degrees of freedom are added in each case. For a much more careful analysis of perturbative correction in fixed-area states see [101]. If we treat the gravitational path integral semiclassically, the Euclidean
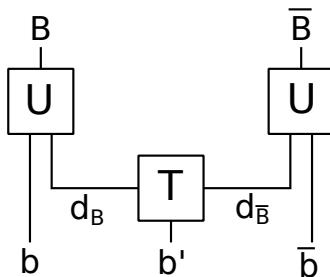
Figure 18. A simple random tensor network, with a single random tensor $|T\rangle$ and three bulk legs. One bulk leg flows into $|T\rangle$, while the other two are mapped directly into the boundary Hilbert spaces B and $\bar{B}$.

field theory path integral in the bottom-half of the geometry prepares a particular state $|\psi\rangle_{bb'\bar{b}}$ of the bulk quantum fields. Here b is the bulk region near B, $\bar{b}$ is the bulk region near $\bar{B}$ and b' is the central region between the extremal surfaces. The field theory partition function on the full replicated geometry then evaluates to

$$\langle\psi|^{\otimes n}\,\tau_b\,\pi_{b'}\,|\psi\rangle^{\otimes n}, \tag{E.8}$$

where the operators τ_b and $\pi_{b'}$ permute the subsystems in regions b and b' respectively.

What about if we add bulk legs to the random tensor network, as shown in figure 18? The answer is exactly the same thing. If we input some bulk state $|\psi\rangle_{bb'\bar{b}}$ and then evaluate $\mathrm{Tr}(\rho_B^n)$ by replacing the random tripartite states $|T\rangle$ by a sum over permutations π, as in (E.3), we find that the term associated to each permutation has an additional factor that is exactly (E.8). The analogy between fixed-area states and random tensor networks continues to hold.

F The Page transition in the simple model

The aim of this appendix is to comprehensively analyse the full Page transition in the simple model. Despite the simplicity of this model, the structure of the transition turns out to be somewhat complicated. Different features have transitions at different times, creating a number of distinct "sub-phases".

As a simple example, the transition happens earlier for larger n Rényi entropies, creating an infinite number of different Page transitions. We will ignore the Rényi entropies, and instead focus only on changes in either

(a) qualitative features of the shape of the entanglement spectrum, or

(b) in the von Neumann entropy.

Nevertheless, we still find seven distinct phases within the transition, as the entanglement spectrum slowly switches from a flat spectrum, with all eigenvalues equal to $1/k$, to a thermal spectrum, with entropy given by the Bekenstein-Hawking entropy of the black hole. Because the details of the calculations in this appendix are fairly technical, we first

summarise the main conclusions about each phase, and then proceed to analysing each phase in detail.

Assumptions and notation. In this appendix, we shall always assume that we are in the semiclassical limit $\beta \ll 1$.[49] At certain points, we shall also assume, for convenience, that the brane mass $\mu \gg 1/\beta$.

We first formalise some notation. Recall that

$$Z_n = e^{S_0} \int ds \rho(s) y(s)^n. \tag{F.1}$$

We use notation where the saddle point value of s for Z_n is denoted by $s^{(n)}$. In the limit $\mu \gg 1/\beta$, we have

$$\rho(s) \, y(s)^n \sim \frac{s}{2\pi^2} \, y(0)^n \, e^{2\pi s - n\beta s^2/2}. \tag{F.2}$$

Hence

$$s^{(n)} = \frac{2\pi}{n\beta} + O(1). \tag{F.3}$$

There is one more saddle point that will be relevant for our calculations. This is the saddle point for

$$e^{S_0} \int ds \rho(s)^2 y(s), \tag{F.4}$$

which we denote by s'. For $\mu \gg 1/\beta$, $s' = 4\pi/\beta$. Note that we have $s' > s^{(1)}$ and $s^{(n)} > s^{(m)}$ for $n < m$.

For a particular choice k, there are two other values of s which will be important. The first, denoted by s_k, is the value of s corresponding to the kth largest thermal eigenstate, as defined in (2.39). In the semiclassical limit, this is defined by

$$\frac{e^{S_0} \rho(s_k)}{2\pi} = k. \tag{F.5}$$

The second, which we denote by $\widetilde{s}_k$ labels the thermal eigenstates with eigenvalue $1/k$. This is defined by

$$\frac{y(\widetilde{s}_k)}{Z_1} = \frac{1}{k}. \tag{F.6}$$

Note that the normalisation of the thermal state ensures that we always have $\widetilde{s}_k < s_k$.

Finally, an important value of k, which we shall denote by $k_{3\to4}$, is defined to be the smallest value of k for which

$$e^{S_0} \rho(\widetilde{s}_{k_{3\to4}}) = \frac{k_{3\to4}^3 \, y(s_{k_{3\to4}})^2}{Z_1^2}. \tag{F.7}$$

For $\mu \gg 1/\beta$, we have $\widetilde{s}_k = \sqrt{2\pi(2s_k/\beta - 2\pi/\beta^2)} + o(1/\beta)$ and hence $s_{k_{3\to4}} = (4-2\sqrt{2})\pi/\beta + o(1/\beta)$. More generally, we expect that $s^{(2)} < s_{k_{3\to4}} < s^{(1)}$. Note that for $k \ll k_{3\to4}$, the left hand side of (F.7) is much smaller than the right hand side.

---

[49]Recall that, in our units, this corresponds to the limit $G_N \to 0$.

Summary of the phase structure. At small k, specifically $k \ll Z_2/y(0)^2$ (phase I), the shape of the entanglement spectrum is a narrow semicircle distribution, of width $4\sqrt{Z_2/kZ_1^2}$, centred on $1/k$. The von Neumann entropy is approximately equal to $\log k$, with a small correction

$$\delta S = -kZ_2/2Z_1^2 \tag{F.8}$$

that comes from the finite width of the semicircle. This is exponentially small with respect to $(S_{\mathrm{BH}} - \log k)$.

For $Z_2/y^2(0) \ll k \ll e^{S_0}\rho(s^{(2)})$ (phase II), the entanglement spectrum is dominated by the same narrow semicircle distribution. The width of this semicircle continues to give the leading correction to the von Neumann entropy. However, there is now a small tail of much larger eigenvalues, that emerge out of the semicircle as k increases. This tail looks like the thermal spectrum, except all the eigenvalues are increase by a constant shift of $\delta\lambda = 1/k$.

When $e^{S_0}\rho(s^{(2)}) \ll k \ll k_{3\to4}$ (phase III), the original semicircle disappears. The spectrum looks like a thermal spectrum, shifted by $\delta\lambda = 1/k$, which is then cutoff after $O(k)$ eigenvalues. We do not have good analytic control over the cutoff region itself. The von Neumann entropy is dominated by eigenvalues near the cutoff, and is still approximately equal to $\log k$. The largest correction to the entropy comes from the width of the cutoff region and has magnitude $O(k^2y(s_k)^2/Z_1^2)$.

When $k_{3\to4} \ll k \lesssim e^{S_0}\rho(s^{(1)})e^{-O(1/\beta)}$ (phase IV), the leading correction instead comes from the eigenvalues in the shifted thermal spectrum, and has size $O(e^{S_0}\rho(\widetilde{s}_k)/k)$. Unlike in phase III, this correction is well controlled analytic; its exact size is given by a simple integral.

The actual Page transition (phase V) for the von Neumann entropy happens at $k \sim e^{S_0}\rho(s^{(1)})$. At this point, the shift in the thermal tail of eigenvalues starts decreasing in order to preserve the normalisation of the state. The shift is now $(1-p_k)/k$, where p_k is the probability of the thermal ensemble being in one of its k largest eigenvalues. The von Neumann entropy is well approximated, up to a correction that peaks at $O(\beta)$ size near the transition, by the entropy of a shifted thermal spectrum, with a hard cutoff after exactly k eigenvalues. This gives a large $O(1/\sqrt{\beta})$ correction compared to the naive Page curve

$$S = \min(\log k, S_{\mathrm{BH}}), \tag{F.9}$$

where S_{BH} is the black hole entropy in the canonical ensemble. This is parametrically larger than the correction for the microcanonical ensemble, because of the fluctuations in the energy.

As k continues to grow, the shift in the thermal part of the spectrum decays, until it becomes neglible for all eigenvalues where the spectrum actually looks thermal. The entanglement spectrum is now the ordinary unshifted thermal spectrum, except that it is smoothly cutoff after k eigenvalues. The von Neumann entropy is approximately equal to the black hole entropy S_{BH}.

Initially, the largest correction to that entropy still comes from the existence of the cutoff (phase VI) and has size

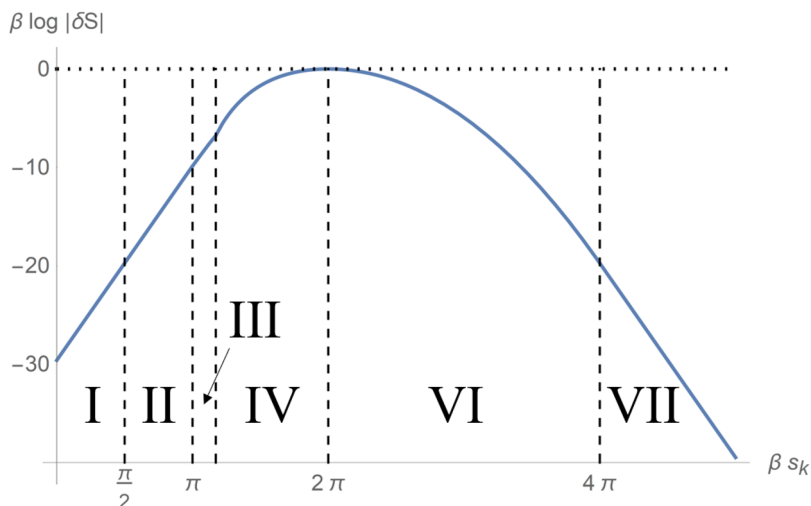$$O(1 - p_k) = O\left(\frac{\rho(s_k)y(s_k)}{Z_1}\right). \tag{F.10}$$

Figure 19. A plot of $\beta \log |\delta S|$ against βs_k in the limit where $\mu \gg 1/\beta$ and $\beta \to 0$. Near the transition, the entropy correction decays as $\delta S = -O(\exp(-\beta(s_k - s^{(1)})^2/2))$ because the largest correction comes from energy fluctuations causing the transition to happen earlier or later than expected. Far from the transition, the entropy decays as $\delta S = -\exp(-O(|s_k - s^{(1)}|))$, as for the microcanonical ensemble. Note that some important features of the correction, such as the maximal size $\delta S = -O(\sqrt{1/\beta})$, are not visible in this plot.

However, when $k \gg e^{S_0}\rho(s')$ (phase VII), this correction is smaller than the effect of small corrections to the thermal spectrum at $s \sim s'$. These corrections give a correction to the entropy of size

$$\delta S = -O\left(\frac{e^{2S_0}\rho(s')^2 y(s')}{kZ_1}\right). \tag{F.11}$$

In the limit $k \to \infty$, we also regain some analytic control over the shape of the spectral cutoff.

Throughout the transition, the von Neumann entropy is well approximated by assuming that the density of states is a thermal spectrum, with a hard cutoff after k eigenvalues, and with the eigenvalues increased by a uniform shift to ensure the correct normalisation. This approximation is worst at the Page transition, when it gives an $O(\beta)$ error.

It is important to note, however, that this shifted thermal spectrum only accurately computes the leading correction to the naive Page curve in phases IV and V. Outside this range, the shifted thermal spectrum entropy is no more accurate than the naive Page curve. Nonetheless, the shifted thermal spectrum is still very accurate everywhere else, since the naive Page curve is itself a very good approximation away from the transition.

In figure 19, we plot the correction δS to the naive Page curve entropy over the course of the Page transition.

Phase I: $k \ll Z_2/y(0)^2$. When k is sufficiently small, we can solve the resolvent equation

$$\lambda R(\lambda) = k + e^{S_0} \int_0^\infty ds \rho(s) \frac{y(s)R}{kZ_1 - y(s)R}, \tag{F.12}$$

by assuming that the resolvent R satsifies $|y(s)R(\lambda)| \ll kZ_1$ for all λ and s. We justify this assumption by checking that the resulting solution for R is self-consistent. We then have

$$\lambda R(\lambda) = k + \int_0^\infty ds \rho(s) \left[\frac{y(s)R}{kZ_1} + \frac{y(s)^2 R^2}{k^2 Z_1^2} + O\left(\frac{y(s^3)R^3}{k^3 Z_1^3} \right) \right], \tag{F.13}$$

$$= k + \frac{R}{k} + \frac{R^2 Z_2}{k^2 Z_1^2}. \tag{F.14}$$

This is a quadratic equation in R with solution

$$R(\lambda) = \frac{-(1/k - \lambda) + \sqrt{(\lambda - 1/k)^2 - 4Z_2/kZ_1^2}}{2Z_2/k^2 Z_1^2}, \tag{F.15}$$

where we chose the correct solution by demanding that $R \to 0$ as $\lambda \to \infty$. The density of states $D(\lambda)$ has support for $1/k - \sqrt{4Z_2/kZ_1^2} \le \lambda \le 1/k + \sqrt{4Z_2/kZ_1^2}$, where it has the form

$$D(\lambda) = \frac{k^2 Z_1^2}{2\pi Z_2} \sqrt{4 \frac{Z_2}{kZ_1^2} - (\lambda - 1/k)^2}. \tag{F.16}$$

This is a semicircle distribution that is sharply peaked around $1/k$.

We now need to check that our assumptions were self-consistent. We have $|R(\lambda)| \lesssim k^{3/2} Z_1 Z_2^{-1/2}$. Hence

$$|R(\lambda)y(s)| \lesssim k^{3/2} Z_1 Z_2^{-1/2} y(0), \tag{F.17}$$

for all s and λ. Our assumption that $|R(\lambda)y(s)| \ll kZ_1$ is therefore valid so long as $k \ll Z_2/y(0)^2$, as claimed.

What about the von Neumann entropy in this phase? At leading order, it is clearly given by $\log k$. To calculate the leading correction to this, we expand λ about $1/k$ to second order. We find

$$S = \int d\lambda D(\lambda)(1/k + (\lambda - 1/k)) \left(\log k - k(\lambda - 1/k) + \frac{k^2(\lambda - 1/k)^2}{2} + O((\lambda - 1/k)^3) \right). \tag{F.18}$$

Using

$$\int d\lambda D(\lambda) = k, \tag{F.19}$$

$$\int d\lambda D(\lambda)(\lambda - 1/k) = 0, \tag{F.20}$$

and

$$\int d\lambda D(\lambda)(\lambda - 1/k)^2 = \frac{Z_2}{Z_1^2}, \tag{F.21}$$

we therefore obtain

$$\delta S = -\frac{kZ_2}{2Z_1^2}. \tag{F.22}$$

This is very similar to the Page result, except that the correction is enhanced by a factor of $Z_2/(Z_1 y(s^{(1)}))$, because the largest correction comes from values of s near $s^{(2)}$ rather than $s^{(1)}$.

It is also worth noting that (F.19) and (F.20) must also be true in the exact solution for $D(\lambda)$, since we know $\text{Tr}(\rho^0) = k$ and $\text{Tr}(\rho) = 1$. This ensures that higher-order perturbative corrections to $D(\lambda)$ give corrections to the von Neumann entropy S that are suppressed compared to (F.22).

A simpler way to reach the same result, without going through the full density of states calculation, is to look at the leading correction to the disconnected geometry in the calculation of the purities $\text{Tr}(\rho^n)$ when k is small. This leading correction comes from geometries where two replicas are connected, with the rest disconnected. There are $n(n-1)/2$ such geometries (from all the ways of pairing replicas), and each geometry is suppressed compared to the disconnected geometry by a factor of $Z_2 k/Z_1^2$. This gives a correction to the Rényi entropy

$$S_n = -\frac{1}{n-1} \log \text{Tr}(\rho^n) \tag{F.23}$$

of size $\delta S_n = -n Z_2 k/2 Z_1^2$. Taking the limit $n \to 1$, reproduces (F.22).

Phase II: $Z_2/y(0)^2 \ll k \ll e^{S_0}\rho(s^{(2)})$. In this regime, the approximation from the previous subsection breaks down at sufficiently small values of s. We need a new approach. Our strategy is to split the integral over s into two pieces: $s < s_{\text{tran}}$ and $s > s_{\text{tran}}$ as

$$\lambda R(\lambda) = k + e^{S_0} \int_0^{s_{\text{tran}}} ds \rho(s) \frac{y(s)R}{kZ_1 - y(s)R} + e^{S_0} \int_{s_{\text{tran}}}^{\infty} ds \rho(s) \frac{y(s)R}{kZ_1 - y(s)R}. \tag{F.24}$$

We treat the integral for $s > s_{\text{tran}}$ in the same way as before. This assumes that

$$|y(s_{\text{tran}})R(\lambda)| \ll kZ_1 \tag{F.25}$$

for all λ. We then treat the integral for $s < s_{\text{tran}}$ as a small perturbation that can be ignored to leading order. For this second approximation to be justified, we require

$$\left| e^{S_0} \int_0^{s_{\text{tran}}} ds \rho(s) \frac{y(s)R}{kZ_1 - y(s)R} \right| \lesssim e^{S_0} \int_0^{s_{\text{tran}}} ds \rho(s) \lesssim e^{S_0} \rho(s_{\text{tran}}) \ll k. \tag{F.26}$$

In the first approximate inequality we used the fact that $|y(s)R| \lesssim |kZ_1 - y(s)R|$, except for a small neighbourhood near the pole at $kZ_1 = y(s)R$. This neighbourhood only gives a small contribution to the integral.

Can we choose s_{tran} to simultaneously satisfy the required conditions on both parts of the integral? The answer is yes. To satisfy our assumption (F.26), we require $s_k - s_{\text{tran}} \gg 1$. In this phase, we have $s_k \ll s^{(2)} \ll s^{(1)}$ and so

$$e^{S_0} \int_{s_{\text{tran}}}^{\infty} ds \rho(s) y(s)^2 \approx Z_2, \qquad e^{S_0} \int_{s_{\text{tran}}}^{\infty} ds \rho(s) y(s) \approx Z_1. \tag{F.27}$$

Treating the second term on the r.h.s. of (F.24) as a small perturbation, we find that the unperturbed solution for the resolvent R is again given by (F.15), as in phase I.

Is this result consistent with our assumption (F.25)? We have

$$|y(s_k)R(\lambda)| \lesssim k^{3/2}Z_1Z_2^{-1/2}y(s_k) = kZ_1\frac{e^{S_0/2}\rho(s_k)^{1/2}y(s_k)}{Z_2^{1/2}} \ll kZ_1, \qquad \text{(F.28)}$$

since $\rho(s_k)y(s_k)^2 \ll \rho(s^{(2)})y(s^{(2)})^2$. We can therefore consistently choose $s_{\text{tran}} = s_k - \kappa$ for some large, but $O(1)$, constant κ, and simultaneously satisfy (F.25) and (F.26).

Having justified our assumptions, we now treat the integral

$$e^{S_0}\int_0^{s_{\text{tran}}} ds\rho(s)\frac{y(s)R}{kZ_1 - y(s)R} \qquad \text{(F.29)}$$

as a small perturbation. In the limit $\lambda - 1/k \gg \sqrt{Z_2/kZ_1^2}$, the unperturbed solution (F.15) reduces to

$$R_0(\lambda) = \frac{k}{\lambda - 1/k}. \qquad \text{(F.30)}$$

Including the perturbation (F.29), we find the first order correction

$$R_1(\lambda) = \frac{e^{S_0}}{\lambda - 1/k}\int_0^{s_{\text{tran}}} ds\rho(s)\frac{y(s)R_0}{kZ_1 - y(s)R_0} = \frac{e^{S_0}}{\lambda - 1/k}\int_0^{s_{\text{tran}}} ds\rho(s)\frac{y(s)/Z_1}{\lambda - 1/k - y(s)/Z_1}. \qquad \text{(F.31)}$$

Since we now have $y(0)/Z_1 \gg \sqrt{Z_2/kZ_1^2}$, the pole at $\lambda - 1/k = y(s)/Z_1$, for sufficiently small s, is in the regime where our approximations are valid. This gives a nonzero density of states, at far larger eigenvalues than anywhere in the semicircle.

Specifically, for $\lambda - 1/k \gg \sqrt{Z_2/kZ_1^2}$, we have

$$D(\lambda) = e^{S_0}\int_0^{s_{\text{tran}}} ds\rho(s)\delta(\lambda - 1/k - y(s)/Z_1). \qquad \text{(F.32)}$$

This looks like a thermal spectrum, with all the eigenvalues shifted upwards by $1/k$. At $\lambda - 1/k \approx 2\sqrt{Z_2/kZ_1^2}$, this shifted thermal spectrum merges into the main semicircle.

The full spectrum is therefore a semicircle, centred on $1/k$ and with width $4\sqrt{Z_2/kZ_1^2}$, plus a very small tail of larger eigenvalues that look like the largest eigenvalues of the thermal spectrum, shifted upwards by $\delta\lambda = 1/k$. Note that there is no tail of small eigenvalues, outside of the semicircle, because in this regime the unperturbed solution is negative, and the perturbed solution does not include any poles.

To understand the exact transition between the shifted thermal tail and the semicircle, we would need to consider the first order perturbation of the full unperturbed solution (F.15). However, we will not do so here, since it is unimportant for our purposes.

We note that the exact solution must satisfy

$$\int_{1/k-\sqrt{4Z_2/kZ_1^2}}^{1/k-\sqrt{4Z_2/kZ_1^2}} d\lambda\, D(\lambda) = k - \int_{1/k+\sqrt{4Z_2/kZ_1^2}}^1 d\lambda\, D(\lambda), \qquad \text{(F.33)}$$

and

$$\int_{1/k-\sqrt{4Z_2/kZ_1^2}}^{1/k-\sqrt{4Z_2/kZ_1^2}} d\lambda\, D(\lambda)(\lambda - 1/k) = -\int_{1/k+\sqrt{4Z_2/kZ_1^2}}^{1} d\lambda\, D(\lambda)(\lambda - 1/k), \qquad \text{(F.34)}$$

because of the constraints $\text{Tr}(\rho^0) = k$ and $\text{Tr}(\rho) = 1$. These results will be crucially to controlling the size of the correction to the von Neumann entropy.

What is the von Neumann entropy? Again, to leading order, it is given by $\log k$. There are two possible sources for the dominant correction. The first is the nonzero width of the semicircle, which gives the same correction that was found in (F.22). The second comes from the existence of the small tail of large eigenvalues. The existence of this tail gives a correction to the entropy of size

$$\delta S_2 = \int_{1/k+\sqrt{4Z_2/kZ_1^2}}^{1} d\lambda D(\lambda)\left[-\lambda\log(\lambda) - \frac{\log k}{k} - (\log k - 1)(\lambda - 1/k)\right]. \qquad \text{(F.35)}$$

The first term here is the direct contribution to the entropy from eigenvalues in the small tail, while the second and third terms come from corrections to the contribution from eigenvalues in the semicircle (calculated by the formula in (F.18)) given (F.33) and (F.34) respectively. We therefore find

$$\delta S_2 = e^{S_0}\int_{1/k+\sqrt{4Z_2/kZ_1^2}}^{1} d\lambda \int_0^{s_{\text{tran}}} ds\,\rho(s)\,\delta(\lambda - 1/k - y(s)/Z_1)(-\lambda\log\lambda - 1/k - \lambda(\log k - 1)) \tag{F.36}$$

$$= e^{S_0}\int_0^{s^*} ds\rho(s)\left[\left(\frac{1}{k} + \frac{y(s)}{Z_1}\right)\left(-\log\left(1 + \frac{ky(s)}{Z_1}\right) + 1\right) - \frac{1}{k}\right], \tag{F.37}$$

$$= \frac{e^{S_0}}{k}\int_0^{s^*} ds\rho(s)\left[\frac{ky(s)}{Z_1} - \left(1 + \frac{ky(s)}{Z_1}\right)\log\left(1 + \frac{ky(s)}{Z_1}\right)\right], \tag{F.38}$$

where $y(s^*) = 2\sqrt{Z_2/k}$ (and hence $s^* \ll s_{\text{tran}}$). Since $0 \geq x - (1+x)\log(1+x) \geq -x^2$ for $x > 0$, we have

$$|\delta S_2| \lesssim \frac{e^{S_0}k\rho(s^*)y(s^*)^2}{kZ_1^2} \ll \frac{kZ_2}{Z_1^2}. \tag{F.39}$$

We conclude that (F.22) is still the leading correction to the von Neumann entropy.

We note that the largest contribution to (F.36) comes from values of $\lambda \sim 1/k + O(\sqrt{Z_2/kZ_1^2})$. In this regime, the approximation $R_0(\lambda) = k/(\lambda - 1/k)$ is not very accurate. Hence the precise size of the correction (F.36) should not be trusted.

This is fine: since (F.19) and (F.20) are fixed by the constraints, the leading effect on the von Neumann entropy of *any* small correction to $D(\lambda)$, with $\lambda - 1/k = O(\sqrt{Z_2/kZ_1^2})$, will be subleading compared to (F.21). It will therefore be small compared to the correction in (F.22). The point of the calculation in (F.36) is to bound the correction from the small density of states with $\lambda - 1/k \gg \sqrt{Z_2/kZ_1^2}$, where a small correction to $D(\lambda)$ could, in principle, have given a larger correction to the von Neumann entropy than (F.22).

Phase III: $e^{S_0}\rho(s^{(2)}) \ll k \ll k_{3\to 4}$. In this phase, we will have somewhat less control over the shape of the peak of the density of states $D(\lambda)$. We will focus on understanding the density of states *away* from this peak. Fortunately this approach will still give us good control over the von Neumann entropy. Our strategy will be similar to our strategy for phase II. However we will find that our approximation for the resolvent $R(\lambda)$ will now only be valid outside of a small range of λ (which includes the peak of the spectrum).

Explicitly, we choose some s_{tran} as for $Z_2/y(0)^2 \ll k \ll e^{S_0}\rho(s^{(2)})$, and approximate the resolvent equation by

$$\lambda R(\lambda) = k + e^{S_0} \int_0^{s_{\text{tran}}} ds\rho(s)\frac{y(s)R}{kZ_1 - y(s)R} + e^{S_0} \int_{s_{\text{tran}}}^\infty ds\rho(s)\frac{y(s)R}{kZ_1}, \tag{F.40}$$

where the second term on the right hand side is treated as a small perturbation. As before, for this approximation to be valid, we require $s_k - s_{\text{tran}} \gg 1$. as well as $|y(s_{\text{tran}})R(\lambda)| \ll kZ_1$, for the values of λ where we want to be able to trust our approximation.

Unlike for phase II, we do not include a term

$$e^{S_0} \int_{s_{\text{tran}}}^\infty ds\rho(s)\frac{y(s)^2 R^2}{k^2 Z_1^2}, \tag{F.41}$$

in our approximation for the integral over $s > s_{\text{tran}}$. In this phase we will have $s_{\text{tran}} \gg s^{(2)}$, causing (F.41) to be dominated by $s \sim s_{\text{tran}}$. It will therefore be much smaller than k, and can be safely ignored, whenever our approximations are themselves valid.

Ignoring the small perturbation, we obtain the initial unperturbed solution

$$R_0(\lambda) = \frac{k}{\lambda - 1}. \tag{F.42}$$

Here, we have used the fact that $s_{\text{tran}} < s_k \ll s^{(1)}$, to see that

$$e^{S_0} \int_{s_{\text{tran}}}^\infty ds\rho(s)y(s) \approx Z_1. \tag{F.43}$$

Our unperturbed solution $R_0(\lambda)$ has a pole at $\lambda = 1/k$, so it is clearly not consistent with our assumption that $|y(s_{\text{tran}})R(\lambda)| \ll kZ_1$ for all values of λ.

What values of λ allow us to choose s_{tran} so that our solution is self-consistent? We have $|y(s_k)R(\lambda)| \ll kZ_1$ so long as

$$|\lambda - 1/k| \gg y(s_k)/Z_1. \tag{F.44}$$

Since $s_k = O(1/\beta)$ in the semiclassical limit, we have $y(s) \propto e^{O(s)}$ for $s \approx s_k$. Both assumptions are therefore valid so long as (F.44) holds and we choose $s_{\text{tran}} = s_k - \kappa$, for some large, but $O(1)$, constant κ. As for phase II, the first order correction to (F.42) is given by (F.31).

We find that the density of states $D(\lambda) = 0$ for sufficiently small λ, satisfying $1/k - \lambda \gg y(s_k)/Z_1$. For large values of λ, with $\lambda - 1/k \gg y(s_k)/Z_1$, we have

$$D(\lambda) = e^{S_0} \int ds\rho(s)\delta(\lambda - 1/k - y(s)/Z_1), \tag{F.45}$$

as in (F.32). This region only contains a small fraction of the k total eigenvalues in the entanglement spectrum — by definition, the kth largest eigenvalue in the shifted thermal spectrum is given by $\lambda = 1/k + y(s_k)/Z_1$. The remaining eigenvalues must lie in the range $1/k - O(y(s_k)/Z_1) \lesssim \lambda \lesssim 1/k + O(y(s_k)/Z_1)$. Since we know that $R_0(\lambda)$ is not a good approximate solution for any values of λ within this range, we conclude that the density of states $D(\lambda)$ is large throughout this region, rather than being concentrated parametrically closer to $1/k$.

In summary, we conclude the following. The density of states looks like a thermal spectrum, shifted by $\delta\lambda = 1/k$, cutoff by some unknown function of width $O(y(s_k)/Z_1)$ at $\lambda = 1/k$. We do not know any compelling way to get analytic traction on this cutoff function, although it can of course be calculated numerically, by solving either the entire resolvent equation numerically or an appropriate approximation to it.

We note that when $k \sim e^{S_0}\rho(s^{(2)})$, the width of the semicircle that we found in phase II is equal to the width of this new uncontrolled cutoff function. There is a smooth transition between the two.

We now move on to calculating the von Neumann entropy. To leading order we still have $S = \log k$. As for phase II, there are two potential sources for the leading correction. The first is a correction from the width of the cutoff, as in (F.18). This is equal to

$$\delta S_1 = -\frac{k}{2}\int_{1/k-O(y(s_k)/Z_1)}^{1/k+O(y(s_k)/Z_1)} d\lambda D(\lambda)(\lambda - \frac{1}{k})^2 = O\left(k^2\frac{y(s_k)^2}{Z_1^2}\right). \tag{F.46}$$

To calculate the exact size of this correction, we would need to use a numerical approximation for the density of states in the cutoff region. The second possible correction comes from the existence of the thermal tail of large eigenvalues. As in (F.36), this is given by

$$\delta S_2 = \frac{e^{S_0}}{k}\int_0^{s_k-O(1)} ds\rho(s)\left[\frac{ky(s)}{Z_1} - \left(1 + \frac{ky(s)}{Z_1}\right)\log\left(1 + \frac{ky(s)}{Z_1}\right)\right]. \tag{F.47}$$

Since we have $s^{(2)} \ll s_k \ll s^{(1)}$, this integral is dominated by values of s with $ky(s) \sim Z_1$. Unlike for $k \ll e^{S_0}\rho(s^{(2)})$, this means that the dominant contribution comes from eigenvalues where the approximation can be trusted. Its size is therefore $O(e^{S_0}\rho(\widetilde{s}_k)/k)$, where $\widetilde{s}_k$ is defined in (F.6).

Comparing the sizes of the two corrections, we have

$$\frac{\delta S_2}{\delta S_1} = O\left(\frac{e^{S_0}\rho(\widetilde{s}_k)Z_1^2}{k^3 y(s_k)^2}\right) \ll 1, \tag{F.48}$$

since we are assuming that $k \ll k_{3\to4}$, for $k_{3\to4}$ defined in (F.7).

We therefore conclude that the dominant correction in this phase comes from the width of the cutoff rather than the thermal tail.

Phase IV: $k_{3\to4} \ll k \lesssim e^{S_0}\rho(s^{(1)})e^{-O(1/\beta)}$. This phase can be dealt with using exactly the same strategy as phase III. The only change is that we now find that the source of the leading correction to the von Neumann entropy (which is still given by $\log k$ at leading order) has changed. Specifically, the leading correction $\delta S_2 = O(e^{S_0}\rho(\widetilde{s}_k)/k)$ now comes

from the existence of the thermal tail, rather than the width of the cutoff region. Its exact size is given (F.47), with high accuracy in the semiclassical limit.

One might worry about whether we can still assume that

$$e^{S_0} \int_{s_{\text{tran}}}^{\infty} ds \rho(s) y(s) \approx Z_1. \tag{F.49}$$

As we shall see in the next section, a more accurate treatment of this integral would decrease the shift in the thermal tail by p_k/k, where p_k is the probability that the thermal ensemble is in one of its k largest eigenvalues. When $k/e^{S_0}\rho(s^{(1)}) = e^{-O(1/\beta)}$, we have $p_k/k = O(y(s_k)/Z_1)$, so this shift can be ignored everywhere that the shifted thermal spectrum can be trusted.

The leading *uncontrolled* correction to the von Neumann entropy still comes from the width of the cutoff region, and has size $\delta S_1 = O(k^2 y(s_k)^2/Z_1^2)$, as in (F.46).

Phase V: $e^{S_0}\rho(s^{(1)})e^{-O(1/\beta)}) \ll k \ll e^{S_0}\rho(s^{(1)})e^{O(1/\beta)}$. We have now finally reached the actual Page transition for the von Neumann entropy! In this phase, we can still use the same approximations that we used in phases III and IV to find the shape of the density of states away from its peak. Again, we want to choose $s_{\text{tran}} = s_k - \kappa$, for a large, but $O(1)$, constant κ. However, we can no longer assume that

$$e^{S_0} \int_{s_{\text{tran}}}^{\infty} ds \rho(s) y(s) = Z_1. \tag{F.50}$$

Instead we have

$$e^{S_0} \int_{s_{\text{tran}}}^{\infty} ds \rho(s) y(s) \approx e^{S_0} \int_{s_k}^{\infty} ds \rho(s) y(s) = (1 - p_k) Z_1, \tag{F.51}$$

where p_k is the probability of the thermal density matrix being in one of its k largest eigenvalues. We therefore find that the density of states $D(\lambda)$ is given by

$$D(\lambda) = \begin{cases} e^{S_0} \int ds \rho(s) \delta\left(\lambda - \frac{1-p_k}{k} - \frac{y(s)}{Z_1}\right) & \lambda - \frac{1-p_k}{k} \gg \frac{y(s_k)}{Z_1}, \\ \text{large and uncontrolled} & \lambda - \frac{1-p_k}{k} = O\left(\frac{y(s_k)}{Z_1}\right), \\ 0 & \frac{1-p_k}{k} - \lambda \gg \frac{y(s_k)}{Z_1}, \end{cases} \tag{F.52}$$

The region $\lambda - (1 - p_k)/k = O(y(s_k)/Z_1)$, which we do not have good analytic control over, contains most of the eigenvalues.

We see that the shift in the thermal part of the spectrum decreases as we move through the Page transition. This can be easily understood as the eigenvalues adjusting to match the twin constraints that $\text{Tr}(\rho) = 1$ and $\text{Tr}(\rho^0) = k$; if we had $O(k)$ thermal eigenvalues, all shifted by $\delta\lambda = 1/k$, we would have $\text{Tr}(\rho) > 1$.

Near $s^{(1)}$, the function $\rho(s)y(s)$ is well approximated by a Gaussian of width $O(\sqrt{1/\beta})$. More precisely, in the limit where the brane mass μ is large, we have

$$\rho(s)y(s) \propto e^{-\frac{\beta}{2}(s-s^{(1)})^2}. \tag{F.53}$$

We therefore have $Z_1 = O(\sqrt{1/\beta}\,\rho(s^{(1)})y(s^{(1)}))$. Also, for $s_k - s^{(1)} \gg \sqrt{1/\beta}$, we have

$$1 - p_k = O\left(\frac{\rho(s_k)y(s_k)}{\sqrt{\beta}(s_k - s^{(1)})\rho(s^{(1)})y(s^{(1)})}\right). \tag{F.54}$$

It follows that

$$\frac{y(s_k)/Z_1}{(1 - p_k)/k} = O\left(\frac{\sqrt{\beta}\rho(s_k)y(s_k)}{(1 - p_k)\rho(s^{(1)})y(s^{(1)})}\right) \ll 1, \tag{F.55}$$

since in this phase we have $s_k - s^{(1)} \ll 1/\beta$. The width of the uncontrolled region is therefore always small compared to the size of a typical eigenvalue in the uncontrolled region.

Let us try to calculate the von Neumann entropy. As a first approximation, we can consider the von Neumann entropy of a thermal spectrum, with a hard cutoff after k eigenvalues, and with each eigenvalue increased by $(1 - p_k)/k$. This agrees with the actual spectrum everywhere that we have control, and obeys the constraints $\mathrm{Tr}\rho^0 = k$ and $\mathrm{Tr}\rho = 1$. The von Neumann entropy of such a state would be given by

$$S = -e^{S_0} \int_0^{s_k} ds\rho(s)\left(\frac{y(s)}{Z_1} + \frac{1 - p_k}{k}\right)\log\left(\frac{y(s)}{Z_1} + \frac{1 - p_k}{k}\right). \tag{F.56}$$

What is the error from this approximation? We have significantly changed the spectrum in the small region where $\lambda - \frac{1-p_k}{k} = O(\frac{y(s_k)}{Z_1})$. To leading order, this part of the spectrum gives a contribution to the von Neumann entropy of $(1 - p_k)\log((1 - p_k)/k)$. Since the zeroth and first moments are fixed by the rest of the spectrum and the constraints $\mathrm{Tr}(\rho) = 1$ and $\mathrm{Tr}(\rho^0) = k$, the leading difference between the shifted thermal spectrum and the actual spectrum will be controlled by

$$\frac{k}{(1-p_k)} \int_{(1-p_k)/k-O(y(s_k)/Z_1)}^{(1-p_k)/k+O(y(s_k)/Z_1)} d\lambda \,\Delta D(\lambda)\left(\lambda - \frac{(1-p_k)}{k}\right)^2 = O\left(\frac{k^2}{(1-p_k)}\frac{y(s_k)^2}{Z_1^2}\right), \tag{F.57}$$

where $\Delta D(\lambda)$ is the diference between the two spectral densities. To find this formula, we expanded λ to second order about $(1 - p_k)/k$. In estimating its size, we used the fact that most of the k eigenvalues are in this region of the spectrum.

How large can this error get? It is largest when $s_k \sim s^{(1)} + O(\sqrt{1/\beta})$. In this case, we have $(1 - p_k) = O(1)$, so the size of the correction is $O(\beta)$.

A cruder approximation is to replace $\log(y(s)/Z_1 + (1 - p_k)/k)$ in (F.56) by $\log\max(y(s)/Z_1, (1 - p_k)/k)$. The largest error from this simplification comes from values of s where $y(s)/Z_1 \sim (1 - p_k)/k$. From such eigenvalues, we have an $O(1)$ error in our estimate of the logarithm. When $s_k \sim s^{(1)}$, this gives a total error of $O(\rho(s)y(s)/Z_1) = O(\sqrt{\beta})$, which is parametrically worse than the shifted thermal spectrum approximation.

How large is the difference between the shifted thermal entropy (F.56) and a naive Page curve given by $S = \min(\log k, S_{\mathrm{BH}})$? Assuming $\mu \gg 1/\beta$,

$$S_{\mathrm{BH}} = \frac{e^{S_0}}{Z_1} \int ds\rho(s)y(s)\log(Z_1/y(s)) = S_0 + \frac{4\pi^2}{\beta} + \frac{1}{2}\log\left(\frac{2\pi}{\beta}\right) + O(1). \tag{F.58}$$

Let $k = Z_1/y(s^{(1)})$, i.e $\widetilde{s}_k = s^{(1)}$. The naive Page curve entropy would be $\log k$. Instead, using the crude approximation discussed above, we find

$$
\begin{aligned}
S &= \frac{e^{S_0}}{Z_1} \int_0^{s^{(1)}} ds \rho(s) y(s) \left[\frac{1}{2}\beta s^2 + \log \frac{Z_1}{y(0)} \right] + \frac{1}{2}\log k + O(1) \\
&= \log k - \sqrt{2\pi\beta} \int_0^{s^{(1)}} ds\, e^{-\frac{\beta}{2}(s-s^{(1)})^2}(s^{(1)} - s) + O(1) \\
&= \log k - \sqrt{\frac{2\pi}{\beta}} + O(1).
\end{aligned}
\tag{F.59}
$$

In the first line, we used the fact that $p_k = 1/2 + o(1)$. In the second line, we used the fact that $\log(Z_1/y(s^{(1)}) = \log k$ and assumed $\mu \gg 1/\beta$.

The leading correction is therefore $O(1/\sqrt{\beta})$; it becomes very large in the semiclassical limit. This is in sharp contrast to the Page curve for Haar random states, or the microcanonical ensemble, where the correction is never larger than $O(1)$. The energy fluctuations in a thermal state parametrically increase the size of the corrections to the naive Page curve.

Phase VI: $e^{S_0}\rho(s^{(1)})e^{O(1/\beta)} \lesssim k \ll e^{S_0}\rho(s')$. In this phase, we have $(1 - p_k)/k = O(y(s_k)/Z_1)$. Hence the approximations made in phase V no longer give us good control over the bottom of the spectrum. Instead, we treat the regimes $\lambda \gg y(s_k)/Z_1$ and $\lambda \ll y(s_k)/Z_1$ separately.

For the former case, we can treat the entire second term in

$$
R(\lambda) = \frac{k}{\lambda} + \frac{e^{S_0}}{\lambda} \int_0^\infty ds \rho(s) \frac{y(s)R}{kZ_1 - y(s)R}
\tag{F.60}
$$

as a small perturbation. Our initial solution is $R_0(\lambda) = k/\lambda$. Hence we have $R_0\, y(s_k) \ll kZ_1$ and our treatment of the second term (F.60) as a small perturbation is self-consistent. We therefore have the first order perturbative correction

$$
R_1(\lambda) = \frac{e^{S_0}}{\lambda} \int_0^\infty ds \rho(s) \frac{y(s)/Z_1}{\lambda - y(s)/Z_1},
\tag{F.61}
$$

which gives

$$
D(\lambda) = e^{S_0} \int ds \rho(s) \delta(\lambda - y(s)/Z_1).
\tag{F.62}
$$

The thermal spectrum is essentially unshifted, everywhere that it is under good analytic control.

For $\lambda \ll y(s_k)/Z_1$, it is easiest to consider λ as a function of negative, real R. Rewriting (F.60), we find

$$
\lambda = \frac{k}{R} - \frac{e^{S_0}}{R} \int_0^\infty ds \rho(s) \frac{y(s)}{y(s) - kZ_1/R}.
\tag{F.63}
$$

For small negative R, the first term dominates, and so λ is negative. For very large negative R, the second term dominates, and so λ is positive, although it approaches zero as $R \to -\infty$.

Note that for negative R, the contribution from the integral is positive for all values of s. At some intermediate value of R, λ should have a maximum. This value will be the bottom of the eigenvalue spectrum.

Suppose we have $R = -\kappa k Z_1/y(s_k)$ for some large, but $O(1)$, constant κ as usual. In this case, the second term dominates and we have

$$\lambda \gg -\frac{k}{R} = \frac{y(s_k)}{\kappa Z_1}. \tag{F.64}$$

We can therefore be confident that $D(\lambda) = 0$ for $\lambda \ll y(s_k)/Z_1$. As for phases III-V, we do not have good control over the shape of the spectrum in the cutoff region, when $\lambda = O(y(s_k)/Z_1)$.

The largest correction to the von Neumann entropy comes from the cutoff region. For eigenvalues in this region, there is an $O(1)$ multiplicative uncertainty in the eigenvalue, which corresponds to an $O(1)$ uncertainty in $\log \lambda$. The total probability of an eigenvalue being in this region is $O(1 - p_k)$. Hence we have

$$S = S_{\text{BH}} - O(1 - p_k) = S_{\text{BH}} - O\left(\frac{e^{S_0}\rho(s_k)y(s_k)}{Z_1}\right), \tag{F.65}$$

where S_{BH} is the entropy of the canonical ensemble. Note that this correction is the same order of magnitude as the difference between S_{BH} and the entropy of the shifted thermal spectrum (F.56). However, we have no good reason to think that the two corrections are the same. The size will be affected by the details of the cutoff, over which we have very little control.

Phase VII: $k \gg e^{S_0}\rho(s')$. In the limit $k \to \infty$, the correction to the thermal entropy from the cutoff region, discussed above, decays as

$$O\left(\frac{\rho(s_k)y(s_k)}{Z_1}\right) = O\left(e^{-\frac{\beta}{2}(s_k - s^{(1)})^2}\right). \tag{F.66}$$

This decays much faster than the largest corrections to the Renyi entropies, which come from planar diagrams that consist of two discs and are suppressed by a factor of $O(1/k)$ compared to the leading connected topology. One can also make general arguments that it is inconsistent with entanglement wedge reconstruction for the correction to decay faster than $O(1/k)$ in this limit [13].

It follows that there should exist some other correction, which cannot come from the cutoff region, that becomes dominant at sufficiently large k. The source of that correction is the second-order perturbative correction to the resolvent. As for phase VI, the unperturbed solution is given by $R_0(\lambda) = k/\lambda$ and the first order perturbative correction $R_1(\lambda)$ is given by (F.61). The second order correction is

$$R_2(\lambda) = \frac{e^{2S_0}Z_1}{k}\int ds_1 ds_2 \frac{\rho(s_1)\rho(s_2)y(s_1)y(s_2)}{(Z_1\lambda - y(s_1))^2(Z_1\lambda - y(s_2))}. \tag{F.67}$$

The second order perturbative correction to the density of states is therefore

$$D_2(\lambda) = \frac{e^{2S_0}}{kZ_1}\int ds_1 ds_2 \frac{\rho(s_1)\rho(s_2)y(s_1)y(s_2)}{y(s_2) - y(s_1)}\delta'(\lambda - y(s_1)/Z_1). \tag{F.68}$$

The pole at $s_1 = s_2$ is dealt with by taking the principal value. The contributions proportional to $\delta(\lambda - y(s_1)/Z_1)$ and $\delta(\lambda - y(s_2)/Z_1)$, which would otherwise exist, cancel under the relabelling $s_1 \leftrightarrow s_2$. We therefore get a correction to the von Neumann entropy of

$$\delta S = -\int_{O(y(s_k)/Z_1)}^1 d\lambda \, D_2(\lambda) \, \lambda \log(\lambda), \tag{F.69}$$

$$= \frac{e^{2S_0}}{kZ_1} \int_0^{s_k - O(1)} ds_1 \int_0^\infty ds_2 \frac{\rho(s_1)\rho(s_2)y(s_1)y(s_2)}{y(s_2) - y(s_1)} [\log(y(s_1)/Z_1) + 1], \tag{F.70}$$

$$= -\frac{e^{2S_0}}{2kZ_1} \int_0^{s_k - O(1)} ds_1 \int_0^{s_k - O(1)} ds_2 \frac{\rho(s_1)\rho(s_2)y(s_1)y(s_2)}{y(s_1) - y(s_2)} \log \frac{y(s_1)}{y(s_2)}. \tag{F.71}$$

In the last step, we dropped the integral over $s_2 > s_k - O(1)$, which is highly suppressed, and then symmetrised the integrand with respect to the relabelling $s_1 \leftrightarrow s_2$. This integral is dominated by values where $s_1 - s_2 = O(1)$. Its magnitude is therefore

$$\delta S = -O\left(\frac{e^{2S_0}}{kZ_1} \int_0^{s_k - O(1)} ds\rho(s)^2 y(s)\right) = -O\left(\frac{e^{2S_0}\rho(s')^2 y(s')}{kZ_1}\right), \tag{F.72}$$

where s' is the saddle point for $\rho(s)^2 y(s)$. This is proportional to $1/k$, as expected. We emphasize that this final result depends crucially on the fact that $k \gg e^{S_0}\rho(s')$. If we still had $k \ll e^{S_0}\rho(s')$, the integral would be dominated by values of s close to s_k and we would have

$$\delta S = -O\left(\frac{e^{S_0}\rho(s_k)y(s_k)}{Z_1}\right). \tag{F.73}$$

This is the same order of magnitude as the correction we found in phase VI. We emphasize that the exact size of the correction still cannot be computed, since the perturbative approximation is poor for $s = s_k - O(1)$.

As with the correction to the entropy in phases I and II, there is a much simpler way to find this entropy correction, by analytically continuing the leading correction to the Renyi entropies. The leading correction to the purities $\text{Tr}(\rho^n)$ in the limit $k \to \infty$ comes from planar diagrams consisting of two connected components. These give a contribution to the purity of

$$\delta \text{Tr}(\rho^n) = \frac{n}{2kZ_1^n} \sum_{p=1}^{n-1} Z_p Z_{n-p} = \frac{ne^{2S_0}}{2kZ_1^n} \int ds_1 ds_2 \rho(s_1)\rho(s_2) \frac{y(s_1)^n y(s_2) - y(s_1)y(s_2)^n}{y(s_1) - y(s_2)}. \tag{F.74}$$

They therefore give a correction to the Rényi entropies $S_n = 1/(1-n) \log(Tr(\rho^n))$ of

$$\delta S_n = -\frac{ne^{2S_0}}{(n-1)2kZ_n} \int ds_1 ds_2 \rho(s_1)\rho(s_2) \frac{y(s_1)^n y(s_2) - y(s_1)y(s_2)^n}{y(s_1) - y(s_2)}. \tag{F.75}$$

In the limit $n \to 1$, this gives the correction to the von Neumann entropy

$$\delta S = -\frac{e^{2S_0}}{2kZ_1} \int ds_1 ds_2 \rho(s_1)\rho(s_2) \frac{y(s_1)y(s_2)}{y(s_1) - y(s_2)} \log \frac{y(s_1)}{y(s_2)}, \tag{F.76}$$

which is exactly what we found before.

The disadvantage of this approach is that it does not know *when* this calculation breaks down. If we only looked at the two connected component topologies, we would expect that the corrections to the Page curve would become large when $k = O(e^{2S_0}\rho(s')^2 y(s')/Z_1)$, which is much larger than the value of k at the Page transition. In contrast, our more careful approach using the resolvent can see that this large correction doesn't exist when $k \ll e^{S_0}\rho(s')$, because the integral in (F.71) is cutoff at s_k.

Finally, in the limit of very large k, we actually regain some increased analytic control over the cutoff region. Specifically, we need $s_k \gg 1/\beta$. We assume for this section that the brane mass $\mu \gg 1/\beta$ and so we have $y(s) = y(0)e^{-\beta s^2/2}$. Suppose we rewrite (F.60) as

$$\lambda R = k - e^{S_0} \int ds \frac{\rho(s)}{-\frac{kZ_1}{y(s)R} + 1}. \tag{F.77}$$

For $s \gg O(1/\beta)$, $y(s)$ is changing much faster than $\rho(s)$. Hence, for sufficiently large, negative, real R, we have

$$e^{S_0} \int ds \frac{\rho(s)}{-\frac{kZ_1}{y(s)R} + 1} \approx e^{S_0} \int_0^{s_R} ds \rho(s), \tag{F.78}$$

where $y(s_R) = -kZ_1/R$ or equivalently

$$s_R = \sqrt{(2/\beta)\log(-Ry(0)/kZ_1)}. \tag{F.79}$$

For self-consistency of our assumptions, we will need $\log(-R\, y(0)/kZ_1) \gg 1$. We know that in the cutoff region we have $|R| = O(kZ_1/y(s_k))$, which implies $s_R \approx s_k$. Our approximation is therefore well controlled so long as $s_k \gg O(1/\beta)$, as claimed.

So far, we have only considered negative, real R. For imaginary or positive R, the integral is seemingly not so simple. However, we can simply deform the integration contour in the complex plane, without passing through any poles, to absorb the phase of R into a phase of $y(s_R)$. Hence (F.79) is valid when R has arbitrary phase so long as we use complex logarithms. Writing $R = -kZ_1 e^{r+i\theta}/y(0)$, we obtain

$$s_R = \sqrt{2r/\beta} + \frac{i\theta}{\sqrt{2\beta r}}, \tag{F.80}$$

where we used the fact that $r \gg 1/\beta$. We can also do the integral in (F.77) explicitly to get

$$-Z_1 \lambda e^{r+i\theta} = 1 - \frac{e^{S_0}}{k} \frac{s_R}{4\pi^3} e^{2\pi s_R}. \tag{F.81}$$

Comparing the imaginary parts of the left and right hand sides, we find

$$Z_1 \lambda e^r \sin\theta = \frac{e^{S_0}}{4\pi^3 k}\sqrt{\frac{2r}{\beta}} e^{2\pi\sqrt{2r/\beta}} \sin\frac{\theta}{\sqrt{2\beta r}} \tag{F.82}$$

Hence, for any fixed $0 < \theta < 2\pi$, we have

$$Z_1 \lambda e^r \ll \frac{e^{S_0}}{k}\sqrt{\frac{2r}{\beta}} e^{2\pi\sqrt{2r/\beta}}. \tag{F.83}$$

Comparing the real parts of (F.81), we have

$$-Z_1 \lambda \, e^r \cos\theta = 1 - \frac{e^{S_0}}{4\pi^3 k}\sqrt{\frac{2r}{\beta}} e^{2\pi\sqrt{2r/\beta}}, \tag{F.84}$$

which, combined with (F.83) tells us that, for fixed $0 < \theta < 2\pi$,

$$1 - \frac{e^{S_0}}{4\pi^3 k}\sqrt{\frac{2r}{\beta}} e^{2\pi\sqrt{2r/\beta}} \ll 1. \tag{F.85}$$

Hence $r - \beta s_k^2/2 = \Delta r \ll \beta s_k$. Substituting this back into (F.82), and using the small angle approximation for $\sin(\theta/\sqrt{\beta r})$, we find

$$\frac{Z_1 \lambda \, e^{\Delta r}}{y(s_k)}\sin\theta = \frac{2\pi\,\theta}{\beta s_k}. \tag{F.86}$$

Hence

$$\theta = \mathrm{sinc}^{-1}\!\left(\frac{2\pi\,y(s_k)}{Z_1\,\lambda\,\beta\,s_k\,e^{\Delta r}}\right), \tag{F.87}$$

where the inverse-sinc function $\mathrm{sinc}^{-1} : [0,1] \to [0,\pi]$ is single-valued on its domain. We now need to solve for Δr. We now turn to (F.84). Given (F.85), we can rewrite (F.84) as

$$\frac{Z_1\lambda e^{\Delta r}}{y(s_k)}\cos\theta = \frac{2\pi\Delta r}{\beta s_k}. \tag{F.88}$$

Solutions for this exist with $\theta = 0$ for $\lambda < 2\pi\, y(s_k)/(e Z_1 \beta s_k) = \lambda_{\min}$, so there are no eigenvalues below this cutoff. At $\lambda = \lambda_{\min}$, we have $\Delta r = 1$.

For $\lambda > \lambda_{\min}$, we have

$$\frac{\lambda\, e^{\Delta r - 1}}{\lambda_{\min}}\cos\mathrm{sinc}^{-1}\!\left(\frac{\lambda_{\min}}{\lambda e^{\Delta r - 1}}\right) = \Delta r. \tag{F.89}$$

In principle, this equation can be solved to find the fixed function $\Delta r(\lambda/\lambda_{\min})$, which does not depend on any parameters. For $\lambda \gg \lambda_{\min}$, we have $\cos(\mathrm{sinc}^{-1}\!\left(\lambda_{\min}e^{-\Delta r + 1}/\lambda\right)) \approx 1$ and so $\Delta r = -W_0(\lambda/\lambda_{\min})$ where W_0 is the Lambert W function. The resolvent R decays at large λ, as expected, and we reenter the regime where our perturbative approximations are well controlled.

Finally, we obtain the density of states

$$D(\lambda) = \frac{Z_1 k e^r \sin\theta}{\pi} = \frac{k Z_1 e^{\Delta r(\lambda/\lambda_{\min})}}{\pi y(s_k)}\sin\mathrm{sinc}^{-1}\!\left(\frac{2\pi\,y(s_k)}{e^{\Delta r(\lambda/\lambda_{\min})}Z_1\lambda\beta s_k}\right). \tag{F.90}$$

For $\lambda \gg y(s_k)/(Z_1\beta s_k)$, we can use the small x approximation $\sin\mathrm{sinc}^{-1}(x) = \pi x$. This gives

$$D(\lambda) = \frac{2\pi\,k}{\lambda\,\beta s_k}. \tag{F.91}$$

For the thermal spectrum near the cutoff, we also have

$$D(\lambda) = e^{S_0}\int_0^{s_k} ds\rho(s)\delta(\lambda - y(s)/Z_1) \approx 2\pi k\int_0^{s_k} ds\,\delta(\lambda - y(s)/Z_1) \approx \frac{2\pi\,k}{\lambda\beta s_k}, \tag{F.92}$$

in the first step, we used $\rho(s) \approx \rho(s_k)$, while the second step used $y'(s) = \beta\,s\,y(s)$. The cutoff spectrum connects smoothly to the thermal spectrum.

This completes our analysis of the entanglement spectrum of the simple model.

References

[1] S.W. Hawking, *Particle Creation by Black Holes*, *[Commun. Math. Phys.](https://doi.org/10.1007/BF02345020)* **43** (1975) 199 [*Erratum ibid.* **46** (1976) 206] [ɪɴSPIRE].

[2] D.N. Page, *Average entropy of a subsystem*, *[Phys. Rev. Lett.](https://doi.org/10.1103/PhysRevLett.71.1291)* **71** (1993) 1291 [[gr-qc/9305007](https://arxiv.org/abs/gr-qc/9305007)] [ɪɴSPIRE].

[3] S. Ryu and T. Takayanagi, *Holographic derivation of entanglement entropy from AdS/CFT*, *[Phys. Rev. Lett.](https://doi.org/10.1103/PhysRevLett.96.181602)* **96** (2006) 181602 [[hep-th/0603001](https://arxiv.org/abs/hep-th/0603001)] [ɪɴSPIRE].

[4] M. Headrick and T. Takayanagi, *A Holographic proof of the strong subadditivity of entanglement entropy*, *[Phys. Rev. D](https://doi.org/10.1103/PhysRevD.76.106013)* **76** (2007) 106013 [[arXiv:0704.3719](https://arxiv.org/abs/0704.3719)] [ɪɴSPIRE].

[5] V.E. Hubeny, M. Rangamani and T. Takayanagi, *A Covariant holographic entanglement entropy proposal*, *[JHEP](https://doi.org/10.1088/1126-6708/2007/07/062)* **07** (2007) 062 [[arXiv:0705.0016](https://arxiv.org/abs/0705.0016)] [ɪɴSPIRE].

[6] A.C. Wall, *Maximin Surfaces, and the Strong Subadditivity of the Covariant Holographic Entanglement Entropy*, *[Class. Quant. Grav.](https://doi.org/10.1088/0264-9381/31/22/225007)* **31** (2014) 225007 [[arXiv:1211.3494](https://arxiv.org/abs/1211.3494)] [ɪɴSPIRE].

[7] A. Lewkowycz and J. Maldacena, *Generalized gravitational entropy*, *[JHEP](https://doi.org/10.1007/JHEP08(2013)090)* **08** (2013) 090 [[arXiv:1304.4926](https://arxiv.org/abs/1304.4926)] [ɪɴSPIRE].

[8] T. Faulkner, A. Lewkowycz and J. Maldacena, *Quantum corrections to holographic entanglement entropy*, *[JHEP](https://doi.org/10.1007/JHEP11(2013)074)* **11** (2013) 074 [[arXiv:1307.2892](https://arxiv.org/abs/1307.2892)] [ɪɴSPIRE].

[9] X. Dong, A. Lewkowycz and M. Rangamani, *Deriving covariant holographic entanglement*, *[JHEP](https://doi.org/10.1007/JHEP11(2016)028)* **11** (2016) 028 [[arXiv:1607.07506](https://arxiv.org/abs/1607.07506)] [ɪɴSPIRE].

[10] M. Rangamani and T. Takayanagi, *Holographic Entanglement Entropy*, *[Lect. Notes Phys.](https://doi.org/10.1007/978-3-319-52573-0)* **931** (2017) 1 [[arXiv:1609.01287](https://arxiv.org/abs/1609.01287)] [ɪɴSPIRE].

[11] N. Engelhardt and A.C. Wall, *Quantum Extremal Surfaces: Holographic Entanglement Entropy beyond the Classical Regime*, *[JHEP](https://doi.org/10.1007/JHEP01(2015)073)* **01** (2015) 073 [[arXiv:1408.3203](https://arxiv.org/abs/1408.3203)] [ɪɴSPIRE].

[12] A. Almheiri, *Holographic Quantum Error Correction and the Projected Black Hole Interior*, [arXiv:1810.02055](https://arxiv.org/abs/1810.02055) [ɪɴSPIRE].

[13] P. Hayden and G. Penington, *Learning the Alpha-bits of Black Holes*, *[JHEP](https://doi.org/10.1007/JHEP12(2019)007)* **12** (2019) 007 [[arXiv:1807.06041](https://arxiv.org/abs/1807.06041)] [ɪɴSPIRE].

[14] G. Penington, *Entanglement Wedge Reconstruction and the Information Paradox*, *[JHEP](https://doi.org/10.1007/JHEP09(2020)002)* **09** (2020) 002 [[arXiv:1905.08255](https://arxiv.org/abs/1905.08255)] [ɪɴSPIRE].

[15] A. Almheiri, N. Engelhardt, D. Marolf and H. Maxfield, *The entropy of bulk quantum fields and the entanglement wedge of an evaporating black hole*, *[JHEP](https://doi.org/10.1007/JHEP12(2019)063)* **12** (2019) 063 [[arXiv:1905.08762](https://arxiv.org/abs/1905.08762)] [ɪɴSPIRE].

[16] A. Almheiri, R. Mahajan, J. Maldacena and Y. Zhao, *The Page curve of Hawking radiation from semiclassical geometry*, *[JHEP](https://doi.org/10.1007/JHEP03(2020)149)* **03** (2020) 149 [[arXiv:1908.10996](https://arxiv.org/abs/1908.10996)] [ɪɴSPIRE].

[17] C. Akers, N. Engelhardt and D. Harlow, *Simple holographic models of black hole evaporation*, *[JHEP](https://doi.org/10.1007/JHEP08(2020)032)* **08** (2020) 032 [[arXiv:1910.00972](https://arxiv.org/abs/1910.00972)] [ɪɴSPIRE].

[18] A. Almheiri, R. Mahajan and J. Maldacena, *Islands outside the horizon*, [arXiv:1910.11077](https://arxiv.org/abs/1910.11077) [ɪɴSPIRE].

[19] M. Rozali, J. Sully, M. Van Raamsdonk, C. Waddell and D. Wakeham, *Information radiation in BCFT models of black holes*, *JHEP* **05** (2020) 004 [arXiv:1910.12836] [ɪɴSPIRE].

[20] H.Z. Chen, Z. Fisher, J. Hernandez, R.C. Myers and S.-M. Ruan, *Information Flow in Black Hole Evaporation*, *JHEP* **03** (2020) 152 [arXiv:1911.03402] [ɪɴSPIRE].

[21] R. Bousso and M. Tomašević, *Unitarity From a Smooth Horizon?*, *Phys. Rev. D* **102** (2020) 106019 [arXiv:1911.06305] [ɪɴSPIRE].

[22] A. Almheiri, R. Mahajan and J.E. Santos, *Entanglement islands in higher dimensions*, *SciPost Phys.* **9** (2020) 001 [arXiv:1911.09666] [ɪɴSPIRE].

[23] X. Dong and A. Lewkowycz, *Entropy, Extremality, Euclidean Variations, and the Equations of Motion*, *JHEP* **01** (2018) 081 [arXiv:1705.08453] [ɪɴSPIRE].

[24] P. Saad, S.H. Shenker and D. Stanford, *A semiclassical ramp in SYK and in gravity*, arXiv:1806.06840 [ɪɴSPIRE].

[25] P. Saad, S.H. Shenker and D. Stanford, *JT gravity as a matrix integral*, arXiv:1903.11115 [ɪɴSPIRE].

[26] P. Saad, *Late Time Correlation Functions, Baby Universes, and ETH in JT Gravity*, arXiv:1910.10311 [ɪɴSPIRE].

[27] J.M. Maldacena, *Eternal black holes in anti-de Sitter*, *JHEP* **04** (2003) 021 [hep-th/0106112] [ɪɴSPIRE].

[28] S.D. Mathur, *The Information paradox: A Pedagogical introduction*, *Class. Quant. Grav.* **26** (2009) 224001 [arXiv:0909.1038] [ɪɴSPIRE].

[29] A. Almheiri, D. Marolf, J. Polchinski and J. Sully, *Black Holes: Complementarity or Firewalls?*, *JHEP* **02** (2013) 062 [arXiv:1207.3123] [ɪɴSPIRE].

[30] R. Bousso, *Complementarity Is Not Enough*, *Phys. Rev. D* **87** (2013) 124023 [arXiv:1207.5192] [ɪɴSPIRE].

[31] Y. Nomura, J. Varela and S.J. Weinberg, *Complementarity Endures: No Firewall for an Infalling Observer*, *JHEP* **03** (2013) 059 [arXiv:1207.6626] [ɪɴSPIRE].

[32] E. Verlinde and H. Verlinde, *Black Hole Entanglement and Quantum Error Correction*, *JHEP* **10** (2013) 107 [arXiv:1211.6913] [ɪɴSPIRE].

[33] K. Papadodimas and S. Raju, *An Infalling Observer in AdS/CFT*, *JHEP* **10** (2013) 212 [arXiv:1211.6767] [ɪɴSPIRE].

[34] J. Maldacena and L. Susskind, *Cool horizons for entangled black holes*, *Fortsch. Phys.* **61** (2013) 781 [arXiv:1306.0533] [ɪɴSPIRE].

[35] M. Headrick, V.E. Hubeny, A. Lawrence and M. Rangamani, *Causality & holographic entanglement entropy*, *JHEP* **12** (2014) 162 [arXiv:1408.6300] [ɪɴSPIRE].

[36] B. Czech, J.L. Karczmarek, F. Nogueira and M. Van Raamsdonk, *The Gravity Dual of a Density Matrix*, *Class. Quant. Grav.* **29** (2012) 155009 [arXiv:1204.1330] [ɪɴSPIRE].

[37] D.L. Jafferis, A. Lewkowycz, J. Maldacena and S.J. Suh, *Relative entropy equals bulk relative entropy*, *JHEP* **06** (2016) 004 [arXiv:1512.06431] [ɪɴSPIRE].

[38] X. Dong, D. Harlow and A.C. Wall, *Reconstruction of Bulk Operators within the Entanglement Wedge in Gauge-Gravity Duality*, *Phys. Rev. Lett.* **117** (2016) 021601 [arXiv:1601.05416] [ɪɴSPIRE].

[39] J. Cotler, P. Hayden, G. Penington, G. Salton, B. Swingle and M. Walter, *Entanglement Wedge Reconstruction via Universal Recovery Channels*, *Phys. Rev. X* **9** (2019) 031011 [arXiv:1704.05839] [INSPIRE].

[40] T. Faulkner and A. Lewkowycz, *Bulk locality from modular flow*, *JHEP* **07** (2017) 151 [arXiv:1704.05464] [INSPIRE].

[41] D. Petz, *Sufficient subalgebras and the relative entropy of states of a von Neumann algebra*, *Commun. Math. Phys.* **105** (1986) 123 [INSPIRE].

[42] D. Petz, *Sufficiency of channels over von Neumann algebras*, *Quart. J. Math. Oxford Ser.* **39** (1988) 97 [INSPIRE].

[43] C.-F. Chen, G. Penington and G. Salton, *Entanglement Wedge Reconstruction using the Petz Map*, *JHEP* **01** (2020) 168 [arXiv:1902.02844] [INSPIRE].

[44] C. Teitelboim, *Gravitation and Hamiltonian Structure in Two Space-Time Dimensions*, *Phys. Lett. B* **126** (1983) 41 [INSPIRE].

[45] R. Jackiw, *Lower Dimensional Gravity*, *Nucl. Phys. B* **252** (1985) 343 [INSPIRE].

[46] A. Almheiri and J. Polchinski, *Models of AdS_2 backreaction and holography*, *JHEP* **11** (2015) 014 [arXiv:1402.6334] [INSPIRE].

[47] P. Hayden and J. Preskill, *Black holes as mirrors: Quantum information in random subsystems*, *JHEP* **09** (2007) 120 [arXiv:0708.4025] [INSPIRE].

[48] J. Maldacena, G.J. Turiaci and Z. Yang, *Two dimensional Nearly de Sitter gravity*, *JHEP* **01** (2021) 139 [arXiv:1904.01911] [INSPIRE].

[49] A. Almheiri, T. Hartman, J. Maldacena, E. Shaghoulian and A. Tajdini, *Replica Wormholes and the Entropy of Hawking Radiation*, *JHEP* **05** (2020) 013 [arXiv:1911.12333] [INSPIRE].

[50] I. Kourkoulou and J. Maldacena, *Pure states in the SYK model and nearly-AdS_2 gravity*, arXiv:1707.02325 [INSPIRE].

[51] D. Harlow and D. Jafferis, *The Factorization Problem in Jackiw-Teitelboim Gravity*, *JHEP* **02** (2020) 177 [arXiv:1804.01081] [INSPIRE].

[52] J. Lin, *Entanglement entropy in Jackiw-Teitelboim Gravity*, arXiv:1807.06575 [INSPIRE].

[53] D.L. Jafferis and D.K. Kolchmeyer, *Entanglement Entropy in Jackiw-Teitelboim Gravity*, arXiv:1911.10663 [INSPIRE].

[54] S.R. Coleman, *Black Holes as Red Herrings: Topological Fluctuations and the Loss of Quantum Coherence*, *Nucl. Phys. B* **307** (1988) 867 [INSPIRE].

[55] S.B. Giddings and A. Strominger, *Loss of Incoherence and Determination of Coupling Constants in Quantum Gravity*, *Nucl. Phys. B* **307** (1988) 854 [INSPIRE].

[56] R. Speicher, *Free probability theory*, arXiv:0911.0087.

[57] P. Cvitanovic, *Planar perturbation expansion*, *Phys. Lett. B* **99** (1981) 49 [INSPIRE].

[58] Z. Yang, *The Quantum Gravity Dynamics of Near Extremal Black Holes*, *JHEP* **05** (2019) 205 [arXiv:1809.08647] [INSPIRE].

[59] A. Kitaev and S.J. Suh, *Statistical mechanics of a two-dimensional black hole*, *JHEP* **05** (2019) 198 [arXiv:1808.07032] [INSPIRE].

[60] C. Akers and P. Rath, *Holographic Renyi Entropy from Quantum Error Correction*, *JHEP* **05** (2019) 052 [arXiv:1811.05171] [INSPIRE].

[61] X. Dong, D. Harlow and D. Marolf, *Flat entanglement spectra in fixed-area states of quantum gravity*, *JHEP* **10** (2019) 240 [arXiv:1811.05382] [INSPIRE].

[62] P. Hayden, R. Jozsa, D. Petz and A. Winter, *Structure of states which satisfy strong subadditivity of quantum entropy with equality*, *Commun. Math. Phys.* **246** (2004) 359.

[63] H. Barnum and E. Knill, *Reversing quantum dynamics with near-optimal quantum and classical fidelity*, *J. Math. Phys.* **43** (2002) 2097.

[64] M. Junge, R. Renner, D. Sutter, M.M. Wilde and A. Winter, *Universal Recovery Maps and Approximate Sufficiency of Quantum Relative Entropy*, *Annales Henri Poincaré* **19** (2018) 2955 [arXiv:1509.07127] [INSPIRE].

[65] D. Harlow and P. Hayden, *Quantum Computation vs. Firewalls*, *JHEP* **06** (2013) 085 [arXiv:1301.4504] [INSPIRE].

[66] A.R. Brown, H. Gharibyan, G. Penington and L. Susskind, *The Python's Lunch: geometric obstructions to decoding Hawking radiation*, *JHEP* **08** (2020) 121 [arXiv:1912.00228] [INSPIRE].

[67] P. Calabrese and J. Cardy, *Entanglement entropy and conformal field theory*, *J. Phys. A* **42** (2009) 504005 [arXiv:0905.4013] [INSPIRE].

[68] S. Sachdev and J. Ye, *Gapless spin fluid ground state in a random, quantum Heisenberg magnet*, *Phys. Rev. Lett.* **70** (1993) 3339 [cond-mat/9212030] [INSPIRE].

[69] A. Kitaev, *A simple model of quantum holography (part 1)*, talk at *KITP*, April 7, 2015, http://online.kitp.ucsb.edu/online/entangled15/kitaev/.

[70] A. Kitaev, *A simple model of quantum holography (part 2)*, talk at *KITP*, May 27, 2015, http://online.kitp.ucsb.edu/online/entangled15/kitaev2/.

[71] A. Kitaev and S.J. Suh, *The soft mode in the Sachdev-Ye-Kitaev model and its gravity dual*, *JHEP* **05** (2018) 183 [arXiv:1711.08467] [INSPIRE].

[72] Y. Gu, A. Lucas and X.-L. Qi, *Spread of entanglement in a Sachdev-Ye-Kitaev chain*, *JHEP* **09** (2017) 120 [arXiv:1708.00871] [INSPIRE].

[73] J.B. Hartle and S.W. Hawking, *Wave Function of the Universe*, *Phys. Rev. D* **28** (1983) 2960 [INSPIRE].

[74] D.N. Page, *Density Matrix of the Universe*, *Phys. Rev. D* **34** (1986) 2267 [INSPIRE].

[75] A.O. Barvinsky, C. Deffayet and A.Y. Kamenshchik, *Density matrix of the Universe reloaded: Origin of inflation and cosmological acceleration*, in *15th International Seminar on High Energy Physics*, (2008) [arXiv:0810.5659] [INSPIRE].

[76] J. Maldacena, *Comments on the Hartle-Hawking wave function of the universe*, talks at *Strings 2019*, Brussels, Belgium, July 9–13, 2019 [https://youtu.be/ZaOb-Z1MhQQ].

[77] J. Cotler, K. Jensen and A. Maloney, *Low-dimensional de Sitter quantum gravity*, *JHEP* **06** (2020) 048 [arXiv:1905.03780] [INSPIRE].

[78] J.M. Maldacena, *Non-Gaussian features of primordial fluctuations in single field inflationary models*, *JHEP* **05** (2003) 013 [astro-ph/0210603] [INSPIRE].

[79] D. Anninos, T. Hartman and A. Strominger, *Higher Spin Realization of the dS/CFT Correspondence*, *Class. Quant. Grav.* **34** (2017) 015009 [arXiv:1108.5735] [INSPIRE].

[80] G.W. Gibbons and S.W. Hawking, *Cosmological Event Horizons, Thermodynamics, and Particle Creation*, *Phys. Rev. D* **15** (1977) 2738 [INSPIRE].

[81] B. Freivogel, Y. Sekino, L. Susskind and C.-P. Yeh, *A Holographic framework for eternal inflation*, *Phys. Rev. D* **74** (2006) 086003 [hep-th/0606204] [INSPIRE].

[82] X. Dong, E. Silverstein and G. Torroba, *de Sitter Holography and Entanglement Entropy*, *JHEP* **07** (2018) 050 [arXiv:1804.08623] [INSPIRE].

[83] A. Lewkowycz, J. Liu, E. Silverstein and G. Torroba, *$T\overline{T}$ and EE, with implications for (A)dS subregion encodings*, *JHEP* **04** (2020) 152 [arXiv:1909.13808] [INSPIRE].

[84] H. Geng, S. Grieninger and A. Karch, *Entropy, Entanglement and Swampland Bounds in DS/dS*, *JHEP* **06** (2019) 105 [arXiv:1904.02170] [INSPIRE].

[85] D. Anninos and D.M. Hofman, *Infrared Realization of dS_2 in AdS_2*, *Class. Quant. Grav.* **35** (2018) 085003 [arXiv:1703.04622] [INSPIRE].

[86] N. Arkani-Hamed, S. Dubovsky, A. Nicolis, E. Trincherini and G. Villadoro, *A Measure of de Sitter entropy and eternal inflation*, *JHEP* **05** (2007) 055 [arXiv:0704.1814] [INSPIRE].

[87] S. Dubovsky, L. Senatore and G. Villadoro, *The Volume of the Universe after Inflation and de Sitter Entropy*, *JHEP* **04** (2009) 118 [arXiv:0812.2246] [INSPIRE].

[88] M.V. Berry, *Semiclassical theory of spectral rigidity*, *Proc. Roy. Soc. Lond. A* **400** (1985) 229.

[89] P. Hayden, S. Nezami, X.-L. Qi, N. Thomas, M. Walter and Z. Yang, *Holographic duality from random tensor networks*, *JHEP* **11** (2016) 009 [arXiv:1601.01694] [INSPIRE].

[90] J.M. Maldacena and L. Maoz, *Wormholes in AdS*, *JHEP* **02** (2004) 053 [hep-th/0401024] [INSPIRE].

[91] S.D. Mathur, *Fuzzballs and the information paradox: A Summary and conjectures*, arXiv:0810.4525 [INSPIRE].

[92] I. Bena and N.P. Warner, *Resolving the Structure of Black Holes: Philosophizing with a Hammer*, arXiv:1311.4538 [INSPIRE].

[93] S. Raju and P. Shrivastava, *Critique of the fuzzball program*, *Phys. Rev. D* **99** (2019) 066009 [arXiv:1804.10616] [INSPIRE].

[94] J.S. Cotler et al., *Black Holes and Random Matrices*, *JHEP* **05** (2017) 118 [*Erratum ibid.* **09** (2018) 002] [arXiv:1611.04650] [INSPIRE].

[95] A. Dabholkar, J. Gomes and S. Murthy, *Nonperturbative black hole entropy and Kloosterman sums*, *JHEP* **03** (2015) 074 [arXiv:1404.0033] [INSPIRE].

[96] J. Maldacena and X.-L. Qi, *Eternal traversable wormhole*, arXiv:1804.00491 [INSPIRE].

[97] A. Strominger, *Les Houches lectures on black holes*, in *NATO Advanced Study Institute: Les Houches Summer School, Session 62: Fluctuating Geometries in Statistical Mechanics and Field Theory*, Les Houches, France, August 2–September 9, 1994 [hep-th/9501071] [INSPIRE].

[98] S. Yakubovich and N. Vieira, *The use of the Kontorovich-Lebedev transform in an analysis of regularized Schrödinger equation*, *Integral Transform. Spec. Funct.* **24** (2013) 9.

[99] F. Pastawski, B. Yoshida, D. Harlow and J. Preskill, *Holographic quantum error-correcting codes: Toy models for the bulk/boundary correspondence*, *JHEP* **06** (2015) 149 [arXiv:1503.06237] [INSPIRE].

[100] C. Akers, T. Faulkner, S. Lin and P. Rath, *Reflected entropy in random tensor networks*, arXiv:2112.09122 [INSPIRE].

[101] X. Dong and D. Marolf, *One-loop universality of holographic codes*, *JHEP* **03** (2020) 191 [arXiv:1910.06329] [INSPIRE].