Entanglement wedge reconstruction and the information paradox

Geoffrey Penington

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

E-mail: geoffp@stanford.edu

ABSTRACT: When absorbing boundary conditions are used to evaporate a black hole in AdS/CFT, we show that there is a phase transition in the location of the quantum Ryu-Takayanagi surface, at precisely the Page time. The new RT surface lies slightly inside the event horizon, at an infalling time approximately the scrambling time $\beta/2\pi \log S_{BH}$ into the past. We can immediately derive the Page curve, using the Ryu-Takayanagi formula, and the Hayden-Preskill decoding criterion, using entanglement wedge reconstruction. Because part of the interior is now encoded in the early Hawking radiation, the decreasing entanglement entropy of the black hole is exactly consistent with the semiclassical bulk entanglement of the late-time Hawking modes, despite the absence of a firewall.

By studying the entanglement wedge of highly mixed states, we can understand the state dependence of the interior reconstructions. A crucial role is played by the existence of tiny, non-perturbative errors in entanglement wedge reconstruction. Directly after the Page time, interior operators can only be reconstructed from the Hawking radiation if the initial state of the black hole is known. As the black hole continues to evaporate, reconstructions become possible that simultaneously work for a large class of initial states. Using similar techniques, we generalise Hayden-Preskill to show how the amount of Hawking radiation required to reconstruct a large diary, thrown into the black hole, depends on both the energy and the entropy of the diary. Finally we argue that, before the evaporation begins, a single, state-independent interior reconstruction exists for any code space of microstates with entropy strictly less than the Bekenstein-Hawking entropy, and show that this is sufficient state dependence to avoid the AMPSS typical-state firewall paradox.

Contents

1 Introduction

By discovering the AdS/CFT correspondence [1, 2], Maldacena definitively answered the question of whether information can escape from a black hole. It can.

While there remains debate about whether information is lost during black hole evaporation in the real universe [3], in AdS/CFT, the bulk quantum gravity theory in $d + 1$ spacetime dimensions is dual to an ordinary d-dimensional conformal field theory that lives on the asymptotic boundary of the bulk spacetime. The unitarity of the boundary conformal field theory means that information must be preserved.

However, on its own, boundary unitarity is not sufficient to consider the information paradox 'solved', even in the restricted context of AdS/CFT. We also need to understand what is *wrong* with the Hawking calculation [4, 5], which apparently suggests that the radiation should be completely thermal until the black hole has almost entirely evaporated,[1] or at least why the conclusion of information loss is naïve.

Conventional effective field theory suggests that the bulk evaporation should be semi-classical, in agreement with Hawking's calculation, so long as the black hole is large compared to the string and Planck scales. In this paper, we will assume that this is indeed the case. In particular, we assume that, from a semiclassical bulk perspective, the Hawking radiation continues to be in a thermal state (up to greybody factors) that is purified by interior modes, even late in the evaporation.

However, as we shall show, this does not mean that no information escapes the black hole. By assuming the quantum version of the Ryu-Takayanagi formula and entanglement wedge reconstruction, we will show that, at late times, the interior degrees of freedom are not microscopically independent of the early Hawking radiation. Instead, a large part of the interior is *encoded in* the early Hawking radiation, in exactly the same way that the bulk in AdS/CFT is microscopically encoded in its asymptotic boundary. This is essentially a formal realisation of the notion of black hole complementarity [6]. We will precisely identify the part of the interior that is encoded in the Hawking radiation, and thereby derive all the expected properties of unitary black hole evaporation.

In particular, we will show that

- Only a non-perturbatively small amount of information escapes the black hole before the so-called Page time, when the entropy of the Hawking radiation becomes equal to the Bekenstein-Hawking entropy of the black hole.[2] However, the existence of such non-perturbatively small corrections is crucial in allowing information to later escape.

- A small diary, thrown into the black hole early in the evaporation, can be reconstructed at the Page time, so long as the state of the black hole is known. A diary thrown into the black hole after the Page time can be reconstructed after waiting for the scrambling time $\beta/2\pi \log S_{BH}$.[3] These twin results are known as the *Hayden-Preskill decoding criterion* and were conjectured based on toy models [8]. We also derive generalisations of the Hayden-Preskill decoding criterion to large diaries and partially unknown black hole states, where we continue to find exact agreement with toy models.

- The microscopic entanglement entropy of the black hole obeys the so-called *Page curve*, which was similarly conjectured based on toy models of black hole evapora-

---

[1]This is a slight over-simplification. In reality, part of the Hawking radiation will be reflected back into the black hole, adding 'greybody factors' to the radiation that escapes.

[2]The Page time is commonly called the 'halfway point' in the black hole evaporation, although, because of the thermodynamic irreversibility of the evaporation and the time dependence of the black hole temperature, it does not occur halfway through the evaporation either by time or by horizon area/entropy [7].

[3]In this formula, S_{BH} is the Bekenstein-Hawking entropy of the black hole, and β is the black hole inverse temperature.

tion [9]. Before the Page time, the entanglement entropy is equal to the entropy of the Hawking radiation, while, after the Page time, it is equal to the Bekenstein-Hawking entropy of the black hole. Crucially, the firewall paradox, which we discuss below, is avoided because the black hole interior is partially encoded in the early Hawking radiation.

The firewall paradox [10] suggests that the combination of standard quantum field theory in the bulk, together with global unitarity, is inconsistent with the existence of a smooth horizon after the Page time. For a smooth horizon with finite energy density to exist, outgoing modes close to the horizon must be entangled with the interior outgoing modes, just inside the horizon. As we evolve forwards in time, these outgoing modes become late-time Hawking radiation.

Furthermore, because the black hole is already close to maximally entangled with the early Hawking radiation, the late-time Hawking radiation must be entangled with the early Hawking radiation, to avoid violating unitarity. However, strong sub-additivity means that a single system cannot be strongly entangled with two different systems at once [11]. We therefore have a paradox. To resolve the paradox, the authors of [10], known by the acronym AMPS, suggested that a 'firewall', or region of very high energy density, must form at the horizon at some point at or before the Page time.

The publication of [10] provoked a flood of responses, including [12–17]. Perhaps most compellingly, in the ER=EPR proposal [18], it was pointed out that the thermofield double state of a CFT,

$$\sum_i e^{-\beta E_i/2} \, |\bar{i}\rangle \, |i\rangle \,, \tag{1.1}$$

is also an example of a black hole that is close to maximally entangled with an external system. Rather than being entangled with the early Hawking radiation, it is entangled with the second copy of the CFT. Hence, the AMPS paradox should apply and the thermofield double state should have a firewall at its horizon. However, the thermofield double state is well-known to be dual to a two-sided Schwarzschild black hole, which has a smooth horizon.

The resolution, in this case, is obvious: the 'interior modes' (which are really exterior modes from the perspective of the second asymptotic boundary) are encoded, from a boundary perspective, in the second copy of the CFT. Hence, there is no contradiction in the Hawking radiation being entangled with both the interior modes and the second copy of the CFT; in fact, the first statement directly implies the second. As one of the results of this paper, we will show that the firewall paradox for one-sided black holes is resolved in exactly the same way.

To show this, and to show all our other results, we will need to use entanglement wedge reconstruction. The entanglement wedge reconstruction conjecture was developed in [19–21] and then established with increasing levels of rigour in [22, 23] and [24]. It has long been known that bulk operators in AdS/CFT can have multiple, distinct boundary representations, which are known as 'reconstructions'. Moreover, there exist reconstructions of any local bulk operator that act on only part of the boundary. Bulk information is encoded redundantly on the boundary. This redundancy is best understood in the lan-

guage of quantum error correction as the statement that bulk operators, acting on the 'code space' of states with a given bulk geometry, are protected against the erasure of certain boundary subregions [25].

Entanglement wedge reconstruction tells us which part of the bulk is encoded in a given part of the boundary. A bulk operator can be reconstructed using a given boundary region B, if, and only if, the bulk operator is contained in a region known as the *entanglement wedge* b of the boundary region B.

To define the entanglement wedge of B, we first need to define the Ryu-Takayanagi surface χ_B associated to B. This surface was originally defined for static spacetimes as the bulk surface χ_B of minimal area $A(\chi_B)$, lying within a static timeslice and homologous to the boundary region B [26, 27]. However, for general dynamic spacetimes [28], and taking into account quantum corrections [29, 30], it is the quantum extremal surface, homologous to B, with the smallest generalised entropy

$$A(\chi_B)/4G_N + S_{\text{bulk}}(\chi_B).$$

Here, the bulk entropy $S_{\text{bulk}}(\chi_B)$ is the von Neumann entropy of the bulk fields contained in the entanglement wedge of B, as defined using the candidate surface, and a quantum extremal surface is defined as a $(d-1)$-dimensional surface of extremal generalised entropy.[4] The Ryu-Takayanagi formula, including quantum corrections, states that entanglement entropy of the boundary region B is equal to the generalised entropy of the Ryu-Takayanagi surface.

The entanglement wedge is now simple to define. It is the bulk region, or, more precisely, the bulk domain of dependence, bounded by the Ryu-Takayanagi surface χ_B and the boundary region B.

An inportant breakthrough was made recently by Almheiri [31], who used entanglement wedge reconstruction to understand how the ER=EPR proposal continues to resolve the firewall paradox for a two-sided black hole, even when the black hole is dynamically evolving in time. He considered a two-dimensional, two-sided black hole, which has an approximate 'boundary' description as an entangled state in a pair of SYK models, and imagined extracting Hawking radiation, using absorbing boundary conditions, from one side of the black hole, and then throwing it into the other side.

When the final state was evolved backwards in time, this time without any interaction between the two sides, he argued that the Ryu-Takayanagi surface was different from the Ryu-Takayanagi surface in the initial state. Degrees of freedom had moved from the entanglement wedge of the 'evaporating' side to the entanglement wedge of the 'growing' side. Information was 'escaping in the Hawking radiation'. Moreover, this change in Ryu-Takayanagi surface meant that the interior modes, with which the Hawking radiation on

---

[4]In this paper, we will always use the term quantum extremal surface to refer to any surface that is an extremum of the generalised entropy. Similarly, classical extremal surface refers to any extremal area surface. We use Ryu-Takayanagi surface, or quantum Ryu-Takayanagi surface, to refer to the quantum extremal surface of minimal generalised entropy and classical Ryu-Takayanagi surface to refer to the minimal area classical extremal surface.

the 'evaporating' side was entangled, were still encoded in the 'growing' side. The two-sided black hole therefore continued to evade the firewall paradox, even as one side of the black hole shrank and the other grew.

The basic conceptual story of this paper will be similar to [31], and indeed to the original ER=EPR proposal [18]. However, rather than relying on the toy model of the thermofield double state, which is well understood but does not actually describe an evaporating black hole, we will work directly with one-sided evaporating black holes.

More specifically, in this paper, we consider an evaporating black hole, formed from collapse, where the Hawking radiation is extracted into an auxiliary reservoir $\mathcal{H}_{\mathrm{rad}}$ using absorbing boundary conditions. By doing so, we will be able to make precise quantitative statements about where, and when, information is encoded.

Unlike in [31], where only classical Ryu-Takayanagi surfaces were considered, it is crucial, when studying an evaporating black hole, that we look at quantum RT surfaces. Since we are extracting the Hawking radiation into an auxiliary reservoir $\mathcal{H}_{\mathrm{rad}}$, the microscopic entanglement entropy of the black hole is simply the entanglement entropy between the entire boundary Hilbert space $\mathcal{H}_{\mathrm{CFT}}$ and the reservoir $\mathcal{H}_{\mathrm{rad}}$. The Ryu-Takayanagi formula states that this entropy is equal to the generalised entropy of the Ryu-Takayanagi surface χ associated to the *entire boundary*.

For an evaporating black hole formed from collapse, the only classical extremal surface, homologous to the entire boundary (i.e. trivial homology), is empty. If entanglement wedge reconstruction was based on the classical Ryu-Takayanagi surface, the interior of the black hole would always be encoded in $\mathcal{H}_{\mathrm{CFT}}$ and no information would ever escape the black hole.[5]

The empty surface is also a quantum extremal surface, with generalised entropy equal to the bulk entanglement entropy S_{rad} between the Hawking radiation and the interior of the black hole. Since we are assuming that the semiclassical Hawking calculation is valid so long as the black hole is large compared to the string/Planck scales, this bulk entanglement entropy will continue to grow, in agreement with semiclassical calculations, even after the Page time.

However, even early in the evaporation of the black hole, it will turn out that there also exists a second, non-empty quantum extremal surface, which lies just inside the event horizon of the black hole. In Eddington-Finkelstein coordinates, the infalling time of this extremal surface is exactly the scrambling time, to leading order, before the 'current time', when Hawking radiation was most recently extracted into $\mathcal{H}_{\mathrm{rad}}$.

Initially, this quantum extremal surface will not be the Ryu-Takayanagi surface. Its generalised entropy will be approximately the Bekenstein-Hawking entropy $S_{\mathrm{BH}} = A_{\mathrm{hor}}/4G_N$ of the black hole, which is much larger than the generalised entropy S_{rad} of the empty surface. However, at the Page time, there will be a phase transition and the non-empty quantum extremal surface will become the Ryu-Takayanagi surface.

---

[5]In this case, we would have to believe either in remnants, or in a complete breakdown of the semiclassical description of the evaporation, as in the firewall proposal. However, remnants are inconsistent with the spectral density of CFTs and there is no reason within the bulk effective field theory to expect the semiclassical bulk description to breakdown until the black hole has almost entirely evaporated.

From this, one can easily use the Ryu-Takayanagi formula to find the entanglement entropy S between the CFT and the reservoir. To leading order, it is given by

$$S = \min(S_{\text{rad}}, A_{\text{hor}}/4G_N). \tag{1.2}$$

The entanglement entropy therefore peaks at the Page time (defined by $S_{\text{rad}} = A_{\text{hor}}/4G_N$) before beginning to decrease. It has long been conjectured that the entanglement entropy of an evaporating black hole is given by this formula, which is known as the Page curve [9]. In particular, a version of the Page curve can be derived if we model the CFT boundary dynamics by a Haar random unitary acting on a large number of qubits.[6] Here we derive it directly from a bulk calculation.[7]

On its own, an explanation of the Page curve using the Ryu-Takayanagi formula is not entirely satisfactory. It does not explain *why* extracting Hawking radiation into the reservoir should decrease the entanglement entropy.

In particular, it does not resolve the firewall paradox. If the entanglement entropy S is to decrease over time, the Hawking radiation that is transferred over from the CFT to the reservoir must itself be entangled with the reservoir. In the semiclassical bulk picture of the evaporation, however, it is instead entangled with the interior of the black hole.

Fortunately, we also know about entanglement wedge reconstruction. The newly emitted Hawking radiation is indeed entangled with interior modes, but some of these modes are now in the entanglement wedge of, and so encoded in, the reservoir $\mathcal{H}_{\text{rad}}$. The same resolution of the firewall paradox that worked for the thermofield double state also works for evaporating black holes.

In the thermofield double state, the Hawking radiation is perfectly thermally entangled with the second CFT. If the late-time Hawking radiation in an evaporating black hole was perfectly thermally entangled with the reservoir, we would find

$$\frac{dS}{dt} = -\frac{dS_{\text{rad}}}{dt} < \frac{1}{4G_N}\frac{dA_{\text{hor}}}{dt}. \tag{1.3}$$

This inequality is strict, even at leading order, because generically black hole evaporation is a strictly thermodynamic-entropy-increasing process. The entanglement structure of the bulk modes would therefore be inconsistent with the Page curve.

However, unlike in the thermofield double state, the Ryu-Takayanagi surface of an evaporating black hole does not lie exactly on the event horizon. Instead it lies an $O(G_N)$ radial distance inside the horizon. Some of the interior outgoing modes are still encoded in the CFT, and so the new radiation is not perfectly entangled with the reservoir $\mathcal{H}_{\text{rad}}$.

In simple cases, one can explicitly calculate the rate of change in the entanglement entropy S that results from extracting a small amount of new Hawking radiation. It agrees

---

[6]This is slightly ahistorical. The Page curve was conjectured well before AdS/CFT was known. However the conjecture was still based on the assumption that the black hole evaporation could be modelled by a Haar random unitary.

[7]We do, of course, need to assume the Ryu-Takayanagi formula and entanglement wedge reconstruction, which are both fundamentally holographic ideas. The Page curve cannot be found using the semiclassical bulk description alone, because it results from the build-up of non-perturbatively small effects. See section 3.3 for more details.

exactly with the rate of change we found using the Ryu-Takayanagi formula. This is not a coincidence. In fact, we shall show that this agreement must always exist; it is a necessary consequence of the Ryu-Takayanagi surface being an extremum of the generalised entropy.

We are also interested in the question of when information about objects thrown into the black hole reappears in the Hawking radiation. If a small diary had been thrown into the black hole more than one scrambling time ago, it would now lie in the entanglement wedge of the reservoir $\mathcal{H}_{\text{rad}}$. It is therefore in principle possible to recover the state of the diary by looking only at the Hawking radiation in the reservoir. At least from a boundary perspective, the information contained in the diary has escaped the black hole.

By modelling the boundary dynamics of the CFT as a fast scrambling unitary, Hayden and Preskill famously conjectured in [8] that the state of a small diary thrown into a black hole early in the evaporation could be decoded from the Hawking radiation at the Page time, while the state of a diary thrown in after the Page time could be decoded after waiting for the scrambling time. Just as for the Page curve, by assuming entanglement wedge reconstruction, we can derive the Hayden-Preskill decoding criterion from a bulk description of the evaporation.

So far, we have avoided any discussion of the crucial issue of state dependence. The idea that there does not exist any single boundary operator that always corresponds to a given interior bulk operator, and instead different boundary operators must be used for different states, goes back to Papadodidimas and Raju [32, 33]. As with the ER=EPR proposal, it was partially inspired as a response to the AMPS firewall paradox. Since then, there has been considerable work on understanding whether such state dependence exists and, if so, how it works [34–36].

In particular, great progress has been made in the context of the SYK model, a toy model of quantum gravity, where it was shown that there exists a complete basis (in fact an overcomplete basis) of pure black hole microstates, whose interior geometries are well understood [37]. Interior operators can be reconstructed on the boundary for each individual microstate, but there is no single reconstruction that works for all the microstates. The idea that the state dependence of interior operators could be interpreted in the language of quantum error correction was suggested in [38] and developed in detail in [31].

In many ways, however, the simplest case of interior state dependence is the Hayden-Preskill decoding criterion. As discussed above, a small diary thrown into a known black hole state can be reconstructed from the Hawking radiation immediately after the Page time. However, to do this, we have to know the state of the black hole.

If there was a way of extracting information about the diary from the Hawking radiation that worked for *any* initial black hole state, then, by linearity, we could also extract information for highly mixed initial black hole states. But for highly mixed intial states, the Hawking radiation will look completely thermal until long after the Page time. So it is clear that the interior operators describing the state of the diary can only be encoded in the Hawking radiation in a highly state-dependent way.

It was shown in [38] that state dependence can arise as a consequence of entanglement wedge reconstruction. By using the formalism of approximate operator algebra quantum error correction, specifically the results of [39–41], one can show that there only exists a

single state-independent reconstruction on a boundary region B of a given bulk operator and for a given code space, if the bulk operator is contained in the entanglement wedge of B for all states, *both pure and mixed*, with support only in the code space. In contrast, the existence of state-dependent reconstructions is possible so long as the bulk operator is contained in the entanglement wedge of B for all *pure states*.

Suppose, as before, we want to reconstruct a small diary, thrown into the black hole at an early time, from the Hawking radiation reservoir $\mathcal{H}_{\mathrm{rad}}$. However, rather than knowing the exact initial state of the black hole, we now only know that the black hole was in some large code space of possible initial microstates. As a simple example, we can imagine that we started with a smaller black hole, in a completely unknown state, and then threw in a large amount of additional energy.

For any pure initial microstate in this code space, the Ryu-Takayanagi surface will jump to the non-empty quantum extremal surface near the horizon, at the Page time, and so the entanglement wedge of $\mathcal{H}_{\mathrm{rad}}$ will contain the diary. If we knew the initial state of the black hole, we could reconstruct the diary. On the other hand, for a highly mixed initial state, the Ryu-Takayanagi surface of the reservoir $\mathcal{H}_{\mathrm{rad}}$ will remain empty until much later.

To be able to find a single state-independent reconstruction that works for the entire code space, we need the interior to be in the entanglement wedge of the reservoir, even for such highly mixed initial states. We therefore need the entropy S_{code} of the code subspace to satisfy

$$S_{\mathrm{code}} < S_{\mathrm{rad}} - S_{BH}. \tag{1.4}$$

Not only does this agree with a conjecture from [38] based on random unitary toy models, it is also provides the mechanism by which information is able to escape the black hole. Regardless of the initial state of the black hole and the state of any diary that was thrown in, the outgoing Hawking radiation is entangled with interior modes in exactly the same way. However, because the interior modes are themselves encoded in the reservoir $\mathcal{H}_{\mathrm{rad}}$ in a state-dependent way, the new Hawking radiation still provides information about the state of the black hole to an observer with access to $\mathcal{H}_{\mathrm{rad}}$.

A similar effect happens before the Page time. At this point, the interior is encoded in the boundary $\mathcal{H}_{\mathrm{CFT}}$ rather than the reservoir $\mathcal{H}_{\mathrm{rad}}$. However the encoding is still necessarily state dependent; if we allow too large a class of initial black hole microstates, the interior will no longer be contained in the entanglement wedge of $\mathcal{H}_{\mathrm{CFT}}$ for highly mixed states. To reconstruct interior operators on the CFT, we need the code subspace of allowed initial microstates to satisfy

$$S_{\mathrm{code}} < S_{\mathrm{BH}} - S_{\mathrm{rad}}. \tag{1.5}$$

If the black hole has not evaporated at all, the bulk entanglement entropy S_{rad} is zero. Hence, (1.5) suggests that we can reconstruct the interior for code spaces of microstates whose entropy is *almost* as large as, but still strictly less than, the Bekenstein-Hawking entropy of the black hole. We will say that the interior of an unevaporated black hole is encoded in $\mathcal{H}_{\mathrm{CFT}}$ in a minimally state-dependent way.

Most of the explicit state-dependent interior reconstructions that have appeared in the literature, for example [32, 33, 36, 37], are only intended to work for a single black hole

microstates, or a code space with $O(1)$ dimension. However, in an appendix, we show that the Kourkoulou-Maldacena construction for the SYK model [37] can be trivially extended to work for a set of microstates with entropy almost as large as the Bekenstein-Hawking entropy. We also show that minimal state dependence is sufficient to avoid the AMPSS typical-state firewall paradox.

The structure of the paper is as follows. In section 2, we study entanglement wedge reconstruction in an evaporating black hole that was formed by collapse. By restricting our attention to a single initial microstate, we avoid the issue of state dependence. We find the location of the non-empty quantum extremal surface explicitly, in a simplified evaporation process where the Hawking radiation is extracted from close to the horizon, in section 2.2, and then use this calculation to explain the Hayden-Preskill decoding criterion and the Page curve in section 2.3. Finally, we show how one can still derive Hayden-Preskill and the Page curve, even when non-trivial greybody factors are present, in section 2.4.

In section 3, we consider large code spaces of initial black hole microstates, and show how the state dependence of interior reconstructions depends on time. We also generalise the Hayden-Preskill decoding criterion to large diaries in section 3.4. In section 3.5, we argue that the interior of black holes that have not evaporated at all is encoded in the boundary with only minimal state dependence.

Finally, section 4 includes a detailed summary of the results of the paper, as well as discussion on various topics. In particular, we argue in section 4.3 that, from a bulk perspective, information must escape the black hole through a version of the Horowitz-Maldacena final state proposal [42]. In appendices, we generalise the calculations from section 2.2 to finite temperature infalling modes, and show how the Kourkoulou-Maldacena construction can easily be made minimally state dependent.

After the completion of this manuscript, the author became aware of independent related work by Almheiri, Engelhardt, Marolf and Maxfield [43], which was published simultaneously.

2 Entanglement wedge reconstruction in an evaporating black hole

In this section, we study an evaporating black hole formed from collapse. For simplicity, we assume throughout that the collapsing matter, and hence the entire spacetime, is rotationally symmetric. We show that no information about the black hole escapes in the Hawking radiation, until the Page time, when the bulk entropy S_{rad} of the Hawking radiation becomes equal to the Bekenstein-Hawking entropy S_{BH} of the black hole. After the Page time, a large part of the interior of the black hole becomes encoded in the early Hawking radiation. In particular, a diary thrown into the black hole becomes encoded in the Hawking radiation after waiting for the scrambling time. The microscopic entanglement entropy of the black hole begins to decrease, in accordance with the Page curve, because the new Hawking radiation is entangled with interior modes that are encoded in the early Hawking radiation.

Our main focus will be on black holes with fixed Schwarzschild radius r_s in AdS units in the limit $G_N \to 0$. All such black holes are microcanonically stable: if we evolve the

system with reflecting boundary conditions, the black hole will quickly reach equilibrium with the Hawking radiation and remain constant in size (up to small fluctuations).[8]

To study the evaporation of microcanonically stable black holes, we instead impose absorbing boundary conditions.[9] The outgoing modes are absorbed by the boundary and so the infalling modes are always in the vacuum state. The Hawking radiation never returns to the black hole, which gradually evaporates.

The dynamics of the system are now irreversible. Rather than evolving unitarily, the boundary state $|\psi\rangle \in \mathcal{H}_{\text{CFT}}$ will obey a Markovian master equation [47]. However, as usual with any quantum channel, we can make the evolution unitary by adding an auxiliary Hilbert space — in this case, a large Markovian reservoir $\mathcal{H}_{\text{rad}}$ that stores the outgoing Hawking radiation once it reaches the boundary. Such a reservoir is sometimes known as an evaporon [45], although we will not use this term. The information paradox will be resolved by simply keeping track of which parts of the bulk are in the entanglement wedge of $\mathcal{H}_{\text{CFT}}$ and which are in the entanglement wedge of $\mathcal{H}_{\text{rad}}$.

We assume that the Ryu-Takayanagi surface, associated to a given boundary region, is defined to be the quantum extremal surface, i.e. surface of extremal generalised entropy

$$\frac{A}{4G_N} + S_{\text{bulk}}, \tag{2.1}$$

homologous to the boundary region, with the smallest generalised entropy. Here, the bulk entropy S_{bulk} is the von Neumann entropy of the bulk fields in any spacelike surface bounded by the Ryu-Takayanagi surface and the boundary region.[10] If the overall bulk state is pure, this bulk von Neumann entropy is simply entanglement entropy between bulk fields inside and outside the entanglement wedge.

We will also generally assume that this prescription is equivalent to a maximin prescription,

$$\max_{\text{Cauchy}} \min_{\chi} \left[\frac{A(\chi)}{4G_N} + S_{\text{bulk}}(\chi) \right], \tag{2.2}$$

where one first finds the surface of (globally) minimal generalised entropy within fixed Cauchy slices, and then selects the Cauchy slice which (globally) maximises this minimal generalised entropy.[11] When this paper first appeared on arXiv, the equivalence of the maximin and extremal (HRT) prescriptions had only been formally shown for classical surfaces (assuming the null energy condition) [21], although it was expected to also be true for quantum surfaces [29]. It has since been shown for quantum surfaces [48] (assuming the

<hr>

[8]This should not be confused with the fact that black holes that are sufficiently small (below the Hawking-Page transition [44]) in AdS units are thermodynamically unstable, even if their size is fixed in the semiclassical limit $G_N \to 0$.

[9]For similar use of absorbing boundary conditions to evaporate black holes in AdS/CFT see [31, 45, 46].

[10]Bulk causality ensures that this bulk entropy is independent of the choice of spacelike surface. Instead S_{bulk} is a function of the bulk domain of dependence of the spacelike surface, which itself depends only on the RT surface and the boundary region.

[11]As just described, the maximin prescription will not generally pick out a unique surface χ, since we can generally find maximising Cauchy slices with more than one minimal surface. However, generically, it should pick out a unique *stable* surface χ, which continues to be close to a minimal surface under any small perturbation of the Cauchy slice. This stable surface should be extremal and the quantum RT surface.
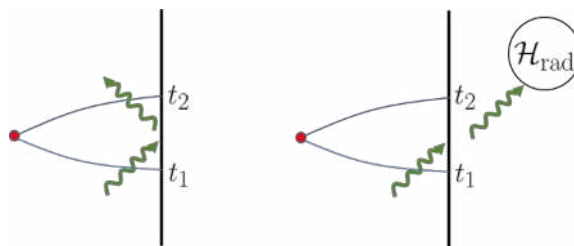
Figure 1. With reflecting boundary conditions, outgoing modes in a surface ending on the boundary at time t_1 may become ingoing modes on a surface ending at time t_2, but the same degrees of freedom will always be contained in each surface. The bulk entropy cannot depend on the boundary time. In contrast, with absorbing boundary conditions, outgoing modes at time t_1 may escape the bulk in the reservoir $\mathcal{H}_{\rm rad}$ by time t_2, and so no longer be contained in a surface ending at time t_2. The bulk entropy, and hence the notion of quantum extremality, depends on the boundary time.

quantum focussing conjecture [49]). We shall therefore only use the maximin prescription to provide intuition about the location of the quantum RT surface; all our actual results will be found using the extremal surface prescription.

If a classical extremal surface is homologous to any boundary component[12] at some time t, it will also be homologous to the same boundary component at any other time and, trivially, will still be a classical extremal surface. The classical Ryu-Takayanagi surface therefore cannot change as a function of the boundary time.[13]

However, this is only true for a quantum extremal surface when we have reflecting boundary conditions. The bulk entropy term in the generalised entropy (2.1) depends not only on local data at the Ryu-Takayanagi surface, but on the state of the bulk fields in entire bulk region, bounded by the RT surface and the boundary.

As shown in figure 1, with reflecting boundary conditions, the same degrees of freedom are contained in this region, independent of the boundary time. However, with absorbing boundary conditions, outgoing modes, which are contained in a spacelike surface that ends on the boundary at time t_1, may escape the boundary and not be contained in a spacelike surface ending at a later time t_2. The bulk entropy, and hence, more importantly, the *gradient* of the bulk entropy, may depend on the boundary time. A quantum extremal surface for the entire boundary at time t_1 may no longer be a quantum extremal surface for the entire boundary at time t_2.

To understand how the information paradox is resolved in AdS/CFT, we will need not only to find the entanglement wedge of the CFT, but also the entanglement wedge of the reservoir $\mathcal{H}_{\rm rad}$. This is not a 'boundary region' in the usual sense and we therefore need to be careful about how to define a Ryu-Takayanagi surface and an entanglement wedge for it.

It is a general fact that, if we divide the boundary of a pure holographic state into two complementary regions, the quantum Ryu-Takayanagi surface will be the same for both regions, and hence the entanglement wedges of the two regions will also be complementary.

---

[12]By this we mean a boundary region B, whose boundary ∂B is empty.

[13]One might worry that the classical Ryu-Takayanagi surface might not be spacelike separated from the boundary at some boundary time. However, this cannot happen, assuming the null energy condition [21].

The same will be true here. The Ryu-Takayanagi surface for $\mathcal{H}_{\mathrm{rad}}$ will be the same as the RT surface for $\mathcal{H}_{\mathrm{CFT}}$, and the entanglement wedge of $\mathcal{H}_{\mathrm{rad}}$ will contain the outgoing modes that were extracted into the reservoir, along with the bulk domain of dependence of any spacelike surface in the black hole interior that is bounded only by the RT surface. All the bulk degrees of freedom will either be in the entanglement wedge of the CFT, or be in the entanglement wedge of the reservoir $\mathcal{H}_{\mathrm{rad}}$.

However, because our 'boundary region' is not actually holographic, let us take a moment to see explicitly why this is true. The simplest way to do so is to imagine throwing the radiation in $\mathcal{H}_{\mathrm{rad}}$ into the bulk of an auxiliary holographic CFT, with a parametrically smaller Newton's constant G'_N (i.e. a parametrically larger central charge) than the original CFT. The small gravitational coupling G'_N ensures that there won't be any significant backreaction.[14] Since the entanglement entropy of this auxiliary CFT must be the same as the entanglement entropy of $\mathcal{H}_{\mathrm{rad}}$, we will define the RT surface of $\mathcal{H}_{\mathrm{rad}}$ to be the RT surface of this auxiliary CFT, which is, by definition the smallest generalised entropy quantum extremal surface homologous to the entire boundary of this auxiliary bulk geometry (i.e. a closed surface with trivial homology). Note that, for a one-sided black hole, the entire boundary of the original CFT also has trivial homology; the two homology constraints are the same.

One might worry that the RT surface that we define in this way could depend on the details of the auxiliary spacetime, in which case it would not be well-defined as an RT surface for $\mathcal{H}_{\mathrm{rad}}$. However, it is easy to see that the RT surface cannot contain a non-empty component in the auxiliary geometry, since in the limit $G'_N \to 0$ the area term will always be dominant and vacuum AdS contains no classical extremal surface.[15] It follows that the only place where a nonemtpy component of the RT surface can exist is in the original black hole geometry.

The entanglement wedge of the auxiliary CFT is the domain of dependence of any spacelike surface bounded by the RT surface and the auxiliary boundary. This will therefore include the entire auxiliary geometry, plus, if the RT surface is non-empty, the domain of dependence of any spacelike region bounded by the RT surface alone.

The RT surface, and entanglement wedge, of the *combination* of a boundary region and a nonholographic system was previously considered in [38]. It was argued there that the RT surface should be the minimal generalised entropy quantum extremal surface, homologous to the boundary region, defined with the nonholographic system automatically included as part of the fields in S_{bulk} (see for example eq. 4.14 of [38]). This was justified by similar arguments to those above (see footnote 15 of [38]). In the special case considered here

---

[14]To avoid talking about more than one holographic CFT, we could instead avoid backreaction by taking a large number of copies of the original CFT and throwing a small amount of radiation into each one.

[15]The quantum extremal surface prescription for the RT surface can be derived from the replica trick [30]. In the limit $G'_N \to 0$, the auxiliary spacetimes will be identical in the replicated and unreplicated geometries (because of the lack of backreaction). It follows that the derivation from [30] can be used to directly calculate the entropy of $\mathcal{H}_{\mathrm{rad}}$, without having to worry about the auxiliary geometry we introduced here. After this paper first appeared on arXiv, these replica trick calculations were done explicitly in [50, 51], where it was shown that the relevant topologies for the second half of the Page curve involve spacetime wormholes connecting different replicas.

where the boundary region is empty and so we only have a nonholographic system, this rule leads to the same conclusions that we reached above.

We have already argued that the RT surfaces for $\mathcal{H}_{CFT}$ and $\mathcal{H}_{\text{rad}}$ satisfy the same (trivial) homology constraint. Moreover, for any given candidate RT surface, the resulting entanglement wedges for $\mathcal{H}_{\text{CFT}}$ and $\mathcal{H}_{\text{rad}}$ are complementary within a Cauchy slice (of the original black hole geometry plus the auxiliary geometry). Since the overall state of the bulk modes is pure, the bulk entropy associated to each entanglement wedge will always be the same. It follows that a quantum extremal surface with respect to $\mathcal{H}_{\text{CFT}}$ will also be a quantum extremal surface with respect to $\mathcal{H}_{\text{rad}}$ and vice versa. The Ryu-Takayanagi surfaces for $\mathcal{H}_{\text{CFT}}$ and $\mathcal{H}_{\text{rad}}$ will therefore be the same.

The simplest quantum extremal surface, for both $\mathcal{H}_{\text{CFT}}$ and $\mathcal{H}_{\text{rad}}$, is, of course, the empty surface. The generalised entropy of the empty surface will be equal to the bulk entanglement entropy S_{rad} between the Hawking radiation and the interior of the black hole. Recall that we assume that the bulk evaporation is semiclassical, so long as the black hole is large compared to the string and Planck scales. Hence the bulk entanglement entropy S_{rad} will continue to grow, in accordance with semiclassical calculations, until the black hole has almost completely evaporated.

Before the Page time, the empty surface will be the Ryu-Takayanagi surface. This can easily be seen using the maximin prescription. Since the empty surface lies in every Cauchy slice, S_{rad} upper bounds the generalised entropy of the Ryu-Takayanagi surface. However, it is also easy to find a Cauchy slice for which the empty surface has minimal generalised entropy. We simply choose a Cauchy slice that stays a small, but fixed, radial distance inside the event horizon of the black hole. Within this Cauchy slice, we cannot choose an RT surface χ that excludes the interior modes, entangled with the outgoing radiation in $\mathcal{H}_{\text{rad}}$, without the area $A(\chi)$ of this surface satisfying

$$\frac{A(\chi)}{4G_N} > S_{\text{rad}}. \tag{2.3}$$

The surface of minimal generalised entropy within this Cauchy slice is therefore the empty surface, with generalised entropy S_{rad}.[16]

It follows that the RT surface is empty and the interior of the black hole is in the entanglement wedge of the boundary CFT, and not the entanglement wedge of the Hawking radiation reservoir. No information has escaped the black hole. The Hawking radiation is thermally entangled with the interior, which is encoded in the CFT, so the redused density matrix of the reservoir $\mathcal{H}_{\text{rad}}$ will be thermal (with appropriate greybody factors). We can also see from the Ryu-Takayanagi formula that the entanglement entropy between the CFT and the reservoir will indeed be the bulk entanglement entropy S_{rad} between the Hawking radiation and the interior. We have derived the first half of the Page curve.

As an aside, we emphasize that, because the evaporation is thermodynamically irreversible, the bulk entanglement entropy S_{rad} is strictly greater than $(A_{\text{hor}}^0 - A_{\text{hor}})/4G_N$

---

[16]Assuming that the minimal generalised entropy surface is unique, it is sufficient, because of the rotational symmetry of the system, to only consider rotational symmetric candidate surfaces. However, it is not difficult to verify that the empty surface does indeed have minimal generalised entropy, within this Cauchy slice, even when we consider surfaces that are not rotationally symmetric.
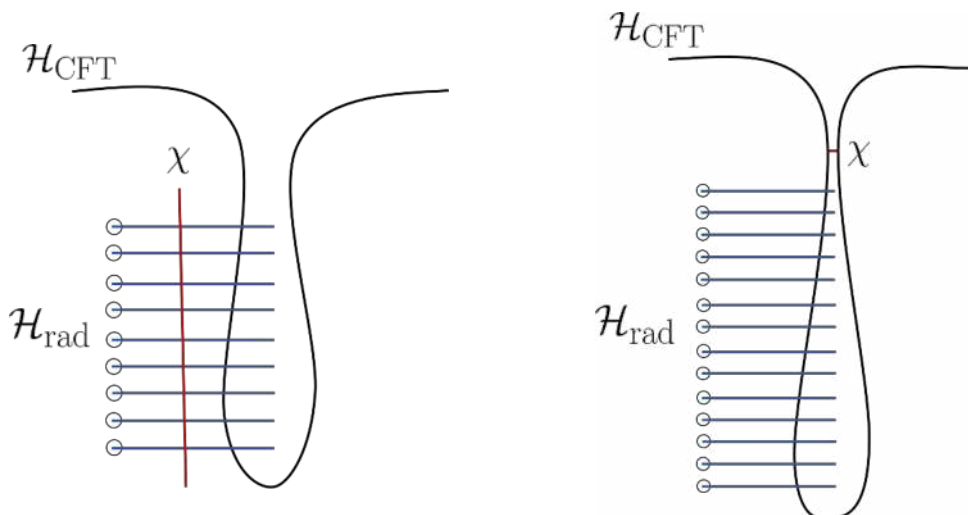
Figure 2. Schematic drawings of Cauchy slices through the black hole interior, both before the Page time (left) and after the Page time (right). The blue lines indicate entanglement between the interior of the black hole and the reservoir $\mathcal{H}_{\text{rad}}$. Before the Page time, there exist Cauchy slices where the empty surface is the surface homologous to the boundary with minimal generalised entropy. It is therefore the Ryu-Takayanagi surface χ. For illustrative purposes, we draw this surface cutting the entanglement between the interior and the reservoir. After the Page time, however, no such Cauchy slice exists. Within any Cauchy slice, there will always exist a surface, near the horizon and homologous to the boundary, with smaller generalised entropy. The Ryu-Takayanagi surface must become non-empty at the Page time.

where A_{hor}^0 is the initial horizon area of the black hole. This means that the Page time, which we recall is defined by

$$S_{\text{rad}} = \frac{A_{\text{hor}}}{4G_N}, \tag{2.4}$$

occurs when the horizon area $A_{\text{hor}} > A_{\text{hor}}^0/2$, despite commonly being called the halfway point of the evaporation [7].

What happens after the Page time? The empty surface is still extremal, but it is easy to see from the maximin prescription that it *cannot* be the Ryu-Takayanagi surface. In any Cauchy slice, we can construct a surface homologous to the boundary that is (a) outside the event horizon, and (b) has only slightly greater area than the current horizon area.[17] Since the only source of greater-than-$O(1)$ bulk entropy is the interior modes that are entangled with radiation that escaped to $\mathcal{H}_{\text{rad}}$, the generalised entropy of this surface will be less than the generalised entropy of the empty surface.[18] The two cases, before and after the Page time, are shown schematically in figure 2.

---

[17]In fact the area of the 'classical maximin surface' that we find in section 2.1 gives an upper bound on this area.

[18]One might worry that late-time Hawking radiation, which has yet to reach the boundary could provide a source of greater-than-$O(1)$ bulk entropy. However, because the Cauchy slice must be achronal, the infalling time of any surface in the Cauchy slice cannot be later than the boundary infalling time. Hence, the bulk entropy of any surface in the exterior will be at most $O(1)$ (once it has been regulated using a cut-off).

We therefore conclude that the quantum Ryu-Takayanagi surface must become non-empty at the Page time. In fact, we will show that there exists a non-empty quantum extremal surface even before the Page time. At the Page time, there is a phase transition, and the non-empty quantum extremal surface becomes the Ryu-Takayanagi surface. The main focus of this section will be on identifying the location of, and the consequences of the location of, this non-empty quantum extremal surface.

2.1 The classical 'maximin surface'

We begin with a warm-up. Since the bulk entropy term $S_{\text{bulk}}(\chi)$ is subleading compared to the area term $A(\chi)/4G_N$ in the formula for the generalised entropy, we shall initially ignore the local variation in the bulk entropy and instead attempt to simply find a 'classical maximin surface' χ_c.[19]

Obviously, for an evaporating black hole formed from collapse, the true classical maximin surface would be empty. We could fix this issue by temporarily assuming that the original black hole was two-sided, with one side allowed to evaporate. However, that would be taking this calculation too seriously. The actual surface that we find will very clearly not make sense as an actual Ryu-Takayanagi surface of any kind (it won't be extremal for example); however, it will turn out to correctly identify the *approximate* location of the quantum extremal surface, which we shall find in the later parts of this section. We shall therefore simply ignore the question of how the black hole formed entirely, and assume that it has been evaporating forever.

Because of the assumed rotational symmetry, the classical maximin surface (and the eventual quantum extremal surface) should be rotationally symmetric. We therefore only need to consider rotationally symmetric Cauchy slices.

We also know that the area of an infalling lightcone in an evaporating black hole is montonically decreasing. Hence, given any Cauchy slice, we can increase

$$\min_{\chi} \frac{A(\chi)}{4G_N} \tag{2.5}$$

by pushing the Cauchy slice backwards and outwards along infalling lightrays. The maximising Cauchy slice is therefore simply the past lightcone of the boundary.

In Eddington-Finkelstein coordinates, the metric of a static black hole is

$$ds^2 = -f(r)dv^2 + 2dvdr + r^2d\Omega^2, \tag{2.6}$$

where the co-ordinate v labels the initial Schwarzschild time of an infalling lightray and, for an uncharged AdS-Schwarzschild black hole,

$$f(r) = 1 + \frac{r^2}{l^2} - \frac{16\pi G_N M}{(d-1)\Omega_{d-1}r^{d-2}}. \tag{2.7}$$

Since our arguments should also be valid for charged black holes (at least when all the particles involved in the Hawking radiation are neutral) and BTZ black holes, we will avoid using (2.7) directly.

---

[19]We specify classical *maximin* surface here, because, unlike the actual quantum extremal surface, this surface will not be extremal, by either the classical or quantum definition.

In the semiclassical limit, the evaporation process is very slow. Over any fixed range of infalling times, the metric will approach the metric of a static black hole of some fixed mass M. We can therefore approximate the metric of an evaporating black hole by a static black hole with a mass M, and hence Schwarzschild radius r_s (defined by $f(r_s) = 0$), that is slowly varying with infalling time v; this is known as an ingoing Vaidya metric.[20]

The radius $r_{l.c.}(v)$ of an outgoing lightcone satisfies

$$\frac{dr_{l.c.}}{dv} = \frac{f(r)}{2} \approx \frac{2\pi}{\beta}(r - r_s), \tag{2.8}$$

where the approximation is valid in the near-horizon region and the inverse temperature $\beta = 4\pi/f'(r_s)$. If $r' = r_{l.c.} - r_s$, we have

$$\frac{dr'_{l.c.}}{dv} = \frac{dr_{l.c.}}{dv} - \frac{dr_s}{dv} \approx \frac{2\pi}{\beta}r'_{l.c.} - \frac{dr_s}{dv}, \tag{2.9}$$

At leading order in the semiclassical limit, we can assume that the inverse temperature β and the evaporation rate dr_s/dv are constant.

Integrating (2.9), we find

$$r_{l.c.} = r_s + C\, e^{\frac{2\pi}{\beta}v} + \frac{\beta}{2\pi}\frac{dr_s}{dv}, \tag{2.10}$$

for some arbitrary constant C. If $C > 0$, the lightcone will eventually escape the black hole, even if its radius is initially decreasing. In contrast, if $C < 0$, the outgoing lightcone will eventually fall into the singularity.

The causal, or event, horizon of the black hole is defined as the boundary of the causal past of future asymptotic infinity. In this case, up to subleading corrections from the time dependence of dr_s/dv and β, it is the outgoing lightcone (2.9) with $C = 0$.[21] Its radius $r_{\rm hor}$ is therefore given by

$$r_{\rm hor} = r_s + \frac{\beta}{2\pi}\frac{dr_s}{dv}. \tag{2.11}$$

Since $dr_s/dv < 0$, this is inside the timelike apparent horizon r_s.[22]

If we instead choose $C = r_s$, we have

$$r_{l.c.} = r_s + r_s\, e^{\frac{2\pi}{\beta}v} + \frac{\beta}{2\pi}\frac{dr_s}{dv} = r_{\rm hor} + r_s\, e^{\frac{2\pi}{\beta}v}, \tag{2.12}$$

and the lightray escapes the near horizon region at $v = O(\beta)$.

---

[20]An ingoing Vaidya metric is only a simple approximation of the actual semiclassical metric of an evaporating black hole; for example, at large radii compared to the Schwarzschild radius, the metric instead resembles an *outgoing* Vaidya metric. For a detailed calculation of the metric of an evaporating black hole in flat space see [52]. In the limit $G_N \to 0$, the only deviation of this metric from the static metric that will be relevant for our calculations is the infalling-time dependence of the black hole mass.

[21]We are assuming here that nothing to radical happens (such as the black hole becoming a white hole) when the black hole has almost entirely evaporated and the semiclassical description breaks down.

[22]For our purposes, the apparent horizon is the radius at which the area of an outgoing lightcone is locally constant.

The radius $r_{l.c.}$, and hence the area $\Omega_{d-1} r^{d-1}$, of the past lightcone reaches a minimum and begins increasing when it reaches the apparent horizon r_s. This occurs at

$$v = -\frac{\beta}{2\pi} \log \frac{r_s}{\beta \, |dr_s/dv|} + O(\beta). \tag{2.13}$$

For small AdS-Schwarzschild black holes,[23] the only relevant lengthscale is the Schwarzschild radius r_s.[24] Hence, it is easy to see using dimensional analysis and the fact that $dr_s/dv = O(G_N)$ that

$$v = -\frac{\beta}{2\pi} \log S_{BH} + O(\beta). \tag{2.14}$$

We have therefore found that the classical maximin surface lies on the classical apparent horizon, one scrambling time into the past.[25]

This is very hopeful: the Hayden-Preskill decoding criterion says that a small diary, thrown into the black hole after the Page time, should be reconstructable from the Hawking radiation after waiting for the scrambling time. Our calculation suggests that this is because the entanglement wedge of the Hawking radiation reservoir $\mathcal{H}_{\text{rad}}$ now contains the diary.

However, there are two major problems with this classical maximin surface χ_c as a candidate Ryu-Takayanagi surface. Firstly, the surface χ_c is not a classical extremal surface. It has extremal area with respect to deformations that stay on the past lightcone, but it certainly does not have extremal area if we allow deformations that move the surface away from the lightcone. It therefore cannot be the Ryu-Takayanagi surface according to the HRT extremal surface prescription, even though the HRT and maximin prescriptions are supposed to be equivalent [21].

Secondly, if the surface χ_c was actually the Ryu-Takayanagi surface, the entanglement wedge of the boundary CFT would not contain the causal future of the boundary. This would be highly problematic because the forward time evolution of the CFT is deterministic, and so the future boundary, although not the past boundary, is in the boundary domain of dependence. The entanglement wedge of the CFT would not contain its causal wedge.

Both problems have the same cause and will have the same solution. The cause is that the evaporating black hole spacetime violates the null energy condition. The null energy

---

[23]For large AdS black holes, there is no clear distinction between radiation that is inside the so-called 'zone' near the horizon and radiation that has escaped to the boundary. As a result, a large AdS black hole does not have a well-defined evaporation rate, even with absorbing boundary conditions; it depends on the details of the evaporation process.

[24]In small, near-extremal Reissner-Nordström black holes, the inverse temperature β is parametrically large compared to the Schwarzschild radius r_s. The approximation in (2.8) is therefore only valid when $r - r_s \ll r_s^2/\beta$. At larger radii, we have $dr/dv = O(1/f(r)) = O(r_s^2/(r - r_s)^2)$ and so an outgoing lightray escapes the black hole in an $O(\beta)$ time. Hence, (2.13) becomes

$$v = -\frac{\beta}{2\pi} \log \frac{r_s^2}{\beta^2 \, |dr_s/dv|} + O(\beta).$$

[25]In addition to the $O(\beta)$ corrections, if the Schwarzschild radius r_s is parametrically small in AdS units, we also need to add the infalling time $\pi \, l_{\text{AdS}}$ for the outgoing lightcone to get from the black hole to the boundary, after it has escaped the near horizon region.

condition is needed to prove that the classical maximin surface is the same as the classical extremal surface [21]. It is also needed to prove that the classical entanglement wedge contains the causal wedge [21].

2.2   The quantum extremal surface

The problem of quantum effects leading to spacetimes that violate the null energy condition was the original reason for the conjecture that the Ryu-Takayanagi surface should be a quantum extremal surface, rather than a classical extremal surface [29]. So, we should definitely be hopeful that all these problems will go away, once we fully include the effects of the bulk entropy term.

Indeed, heuristically, it is easy to see that the bulk entropy term can push the Ryu-Takayanagi surface away from the past lightcone. If the RT surface was exactly on the past lightcone, no outgoing modes would be included in the entanglement wedge of the CFT. Since the entropy of the outgoing modes is divergent, moving the RT surface a small distance inside the lightcone should increase the bulk entropy by a formally infinite amount. This strongly suggests that the quantum RT surface, in the maximin prescription, will be stabilised a small radial distance away from the past lightcone, creating an actual quantum extremal surface.

Unfortunately, actually calculating the bulk entropy is complicated by the presence of non-trivial greybody factors. Because the black hole spacetime is curved, outgoing modes close to the black hole do not necessarily escape to infinity. Instead, there is a non-trivial scattering process. The curved spacetime wave equation can be rewritten as a flat space wave equation with a potential barrier. This potential barrier lies an $O(r_s)$ distance away from the black hole horizon and is higher for modes with large angular momentum, causing the Hawking radiation to be dominated by modes with $O(1)$ angular momentum. The region inside the potential barrier is known as the zone.

Within the zone, the Hawking radiation is truly black body radiation at the black hole temperature.[26] However, the probability of a Hawking mode escaping depends on its angular momentum and, importantly, on its Schwarzschild energy. This probability is known as a greybody factor.

Outgoing modes in the entanglement wedge of the CFT are entangled both with modes further in towards the black hole, and with outgoing modes further out — outside of the past lightcone. The non-trivial greybody factors mean that the outgoing modes outside the past lightcone are related to later *ingoing* modes, which are also in the entanglement wedge of the CFT. This dramatically complicates any explicit calculation of the bulk entropy.

As a simple solution to this problem, we shall therefore temporarily assume that the Hawking radiation is extracted from deep inside the zone, close to the horizon, before the

---

[26]A more precise statement is that each angular momentum mode looks like two-dimensional black body radiation. For higher-dimensional black bodies, only angular momentum modes with $|J| \lesssim Tr$ are significantly excited, and only modes with $|J| \ll Tr$ look like two-dimensional thermal radiation. In contrast, sufficiently close to the horizon, Rindler modes with $|J| \gg Tr_s = O(1)$ will be excited.

mixing of ingoing and outgoing modes occurs.[27] (We will reintroduce the greybody factors in section 2.4.) This obviously involves changing the dynamics of the system compared to our original procedure for extracting the Hawking radiation. In particular, the boundary Master equation, and its purification using the reservoir $\mathcal{H}_{\text{rad}}$, will now involve reconstructions of operators deep in the bulk, which will be highly non-local from a boundary perspective.

We emphasize, however, that there is nothing fundamentally unphysical about this. The outgoing modes will still be extracted at a distance from the horizon (and hence an energy scale) that is fixed in AdS units as $G_N \to 0$; the distance simply needs to be small compared to the Schwarzschild radius r_s. We are therefore still well within the domain of validity of the bulk effective field theory.

It is important to note that the closer to the horizon we extract the Hawking radiation, the larger the number of angular momentum modes that are excited.[28] Similarly, increasingly massive fields will be excited very close to the horizon. This can be seen from the explicit form of the potential barrier in tortoise coordinates [54].

We will assume that we extract some fixed finite number of angular momentum modes for each field, and that these fields are extracted sufficiently close to the horizon that their mass and angular momentum can be safely ignored. In effect, we are changing the dynamics of the theory such that all greybody factors are either zero or one, depending on the angular momentum. We will also assume that all the light fields are free.

Close to the horizon, the spacetime can be approximated by $\mathbb{R}^{1,1} \times S^{d-1}$ where the radius of the sphere S^{d-1} is the Schwarschild radius r_s. Each angular momentum mode acts as an independent free field in an effective two-dimensional theory, with a Kaluza-Klein mass $m_{KK}^2 = L^2/r_s^2$, which can be ignored at the lengthscales of interest.

Let the number of two-dimensional bosonic modes N_b and fermionic modes N_f. The $(1+1)$-dimensional Stefan-Boltzman law [55] states that the rate of energy loss from the black hole is given by

$$\frac{dM}{dv} = \frac{c_{\text{evap}}\,\pi}{12\,\beta^2}, \tag{2.15}$$

where $c_{\text{evap}} = N_b + N_f/2$.[29] The first law of black hole thermodynamics says that $\beta dM = dA_{\text{hor}}/4G_N$. Hence

$$\frac{dA_{\text{hor}}}{dv} = \frac{c_{\text{evap}}\,\pi\,G_N}{3\,\beta}, \tag{2.16}$$

---

[27]By doing so, we are effectively reducing the problem to a calculation in the two-dimensional effective theory that governs the near horizon region [53]. For this reason, our results in this section will be identical (once constants are fixed appropriately) to those found in an explicit two-dimensional model in [43].

[28]Another way of saying this is that modes with larger angular momentum are reflected back into the black hole at a smaller, but still finite, distance from the horizon.

[29]Of course, since Hawking radiation is stochastic, this is only the average rate of energy loss. However, so long as we consider timescales that are large compared to the thermal time β, the average energy change should be large compared to the fluctuations in the energy change, which can therefore be safely ignored. In our case, the relevent timescale is the scrambling time, which is indeed very large compared to the thermal time β in the semiclassical limit. We can also suppress the fluctuations by taking the limit where c_{evap} is large.

and

$$\frac{dr_s}{dv} = \frac{c_{\text{evap}}\,\pi\,G_N}{3\,\beta\,(d-1)\,r_s^{d-2}\,\Omega_{d-1}}.$$

(2.17)

Substituting (2.17) into (2.13) and dropping $O(\beta)$ terms, we find that the classical maximin surface for this spacetime occurs at[30]

$$v_c = -\frac{\beta}{2\pi}\log\frac{S_{BH}}{c_{\text{evap}}} + O(\beta).$$

(2.18)

To calculate the quantum extremal surface, we also need to calculate how the bulk entanglement entropy depends on the location of the extremal surface.[31] Since we are assuming that all the relevant fields are free and effectively massless at the lengthscales of interest, the ingoing and outgoing modes are decoupled. The total bulk entropy is therefore simply the sum of the bulk entropies of the ingoing and outgoing modes.

The infalling modes are in the vacuum state with respect to the infalling time v. Moreover, if we assume (correctly) that the quantum extremal surface is close to the classical maximin surface, the entanglement wedge of the boundary CFT will contain infalling modes spread over approximately the scrambling time, which diverges as $G_N \to 0$. Since the entanglement entropy of the vacuum state grows only logarithmically with system size, we can treat the entanglement entropy of the infalling modes as approximately independent of the location of the extremal surface, so long as the cut-off at the extremal surface is fixed in units of infalling time v. By this, we really mean that the cut-off is equal to $\varepsilon\partial/\partial v$ for some constant ε.

What about the outgoing modes? The only relevant modes are the modes that are extracted into the reservoir $\mathcal{H}_{\text{rad}}$. At sufficiently short lengthscales, the entropy of these modes will be given by the Minkowski vacuum formula [57, 58]

$$S = \frac{c_{\text{evap}}}{6}\log\left(\frac{r_{lc}(v) - r}{\sqrt{\varepsilon_1\varepsilon_2}}\right),$$

(2.19)

---

[30]As noted in Footnote 24, for near-extremal black holes, (2.13) becomes

$$v_c = -\frac{\beta}{2\pi}\log\frac{r_s^2}{\beta^2\,|dr_s/dv|} + O(\beta).$$

Hence, substituting (2.17), we find that

$$v_c = -\frac{\beta}{2\pi}\log\Delta S_{BH} + O(\beta),$$

where $\Delta S_{BH} = S_{BH} - S_{BH}^0 = O(r_s S_{BH}/\beta)$ with S_{BH}^0 the entropy of an extremal black hole with the same charge. The location we find for the non-empty extremal surface is therefore consistent the similar calculations, for two-sided black holes in JT gravity, done in [43]. The $O(\beta)$ corrections are also the same in both calculations [56].

[31]Since the entanglement entropy of gravitons is not well understood, we shall assume here that no graviton modes are extracted into the reservoir $\mathcal{H}_{\text{rad}}$. One would hope that, if we did understand the entanglement entropy of gravitons, we would find that they would contribute to the location of the quantum extremal surface in a similar way to other bulk modes.
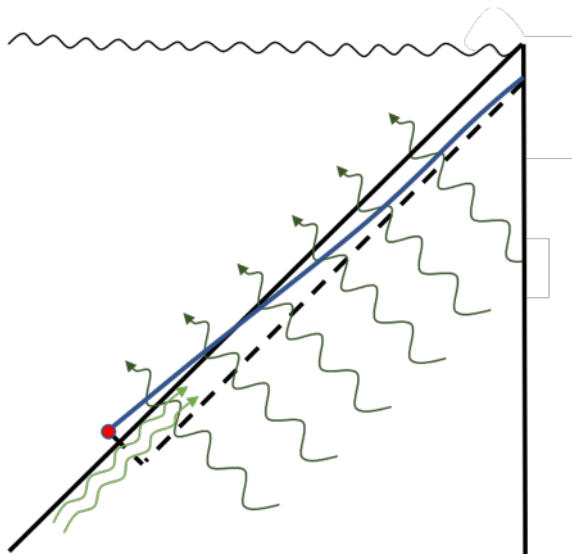
Figure 3. The bulk entropy of the region, shown in blue, between the Ryu-Takayanagi surface and the boundary can be decomposed into the entropy of the ingoing and outgoing modes. The ingoing modes are in the infalling vacuum, and the bulk region includes ingoing modes spread over approximately the scrambling time, which diverges in the semiclassical limit. This means that the gradient, in units of infalling time, of the entropy of the infalling modes tends to zero. In contrast, as the RT surface approaches the past lightcone of the boundary, there will be a negative logarithmic divergence in the (renormalised) entropy of the outgoing modes. This divergence should stabilise the location of the quantum extremal surface a small distance away from the outgoing lightcone.

where r_{lc} is the radius of the outgoing lightcone and ε_1 and ε_2 are the cut-offs at the quantum extremal surface and the outgoing lightcone respectively, in units of the radius r.[32]

The cut-offs ε_1 and ε_2 are crucial to the calculation and so we take some time to discuss them in detail. The cut-off ε_1 at the quantum extremal surface is unphysical. Since the bulk entropy would otherwise be formally divergent, we need to cut-off the bulk degrees of freedom at some fixed proper lengthscale. The lengthscale chosen is arbitrary; in the calculation of physical quantities, such as entanglement entropies in the boundary CFT, the cut-off dependence should be cancelled by the scale dependence of the couplings in the effective gravitational theory due to renormalisation.[33]

Of course, the cut-off ε_1 is not itself a proper lengthscale. Since the radius r is a null coordinate, the proper length of the cut-off ε_1 is zero. The proper cut-off $\varepsilon_{\text{prop}}$ is instead determined by the *inner product* of the cut-off ε_1 with the cut-off at the extremal surface

---

[32]Here we are implicitly using the fact that the radius r, like any smooth, non-singular coordinate, is an approximate affine coordinate along ingoing lightrays at sufficiently small distance scales.

[33]Of course, even boundary entanglement entropies are not actually well defined because of UV-divergences in the boundary theory (which correspond to IR-divergences in the bulk theory). However the mutual information between two boundary regions, for example, is a well defined regulator-independent finite quantity.

on the *ingoing* modes. Specifically

$$\varepsilon_{\text{prop}} = \sqrt{g\left(\varepsilon_1 \frac{\partial}{\partial r}, \varepsilon_{\text{in}} \frac{\partial}{\partial v}\right)}, \tag{2.20}$$

where g is the metric and ε_{in} is the cut-off on the ingoing modes in units of v.

When we claimed that the entropy of the ingoing modes was approximately constant, we implicitly assumed that the cut-off ε_{in} was constant. Since, to leading order, the metric is given by

$$ds^2 = 2dvdr, \tag{2.21}$$

everywhere in the near horizon region, this means that the cut-off ε_1 should also be constant, so that the proper cut-off $\varepsilon_{\text{prop}}$ is constant in AdS units.

The physical status of the lightcone cut-off ε_2 is very different. The lightcone cut-off ε_2 is related to the cut-off ε_0 on the Schwarzschild frequency of the modes that we extract into the reservoir $\mathcal{H}_{\text{rad}}$. Unlike the cut-off at the extremal surface, this is a physical cut-off that depends on the dynamics that we use to extract outgoing modes into $\mathcal{H}_{\text{rad}}$.

However the cut-off ε_2 is blueshifted as the outgoing modes evolve back in time. If we parallel transport the cut-off $\varepsilon_2 \partial/\partial r$ backwards along the past lightcone, we find that

$$0 = \nabla_v \left(\varepsilon_2 \frac{\partial}{\partial r}\right), \tag{2.22}$$

$$= [\partial_v \varepsilon_2 + \Gamma^r{}_{vr} \varepsilon_2] \frac{\partial}{\partial r}, \tag{2.23}$$

$$= \left[\partial_v \varepsilon_2 - \frac{2\pi}{\beta} \varepsilon_2\right] \frac{\partial}{\partial r}, \tag{2.24}$$

where we approximated the metric by its leading order (static) approximation (2.6) and used the fact that, in the near horizon region, $f'(r) \approx f'(r_s) = 4\pi/\beta$. We therefore find that the cut-off ε_2 is related to the (constant) cut-off ε_0 on the Schwarzschild energy of the extracted modes by

$$\varepsilon_2 \propto e^{\frac{2\pi}{\beta}v} \varepsilon_0. \tag{2.25}$$

We emphasize that the infalling-time dependence of this cut-off is purely a product of the coordinate system that we are using. In section 2.4, we do a more general calculation, which includes greybody factors, in Kruskal-Szekeres-like coordinates, where there is no blueshifting, and so the cut-off at the lightcone would be constant.[34] The calculations done here are rederived as a special case, without any reference to exponentially small cut-offs. For the moment, however, we shall continue to use Eddington-Finkelstein coordinates, which have a more natural physical interpretation. For pedagogical purposes, in appendix A, we also give an example of a simple Rindler space calculation that illustrates the importance of taking into account the coordinate dependence of cut-offs.

---

[34]In many ways, the nicest coordinates to use for the problem are the outgoing Kruskal coordinate U together with the infalling time v. However, the author is too lazy to rewrite all the calculations in these coordinates. See [59].
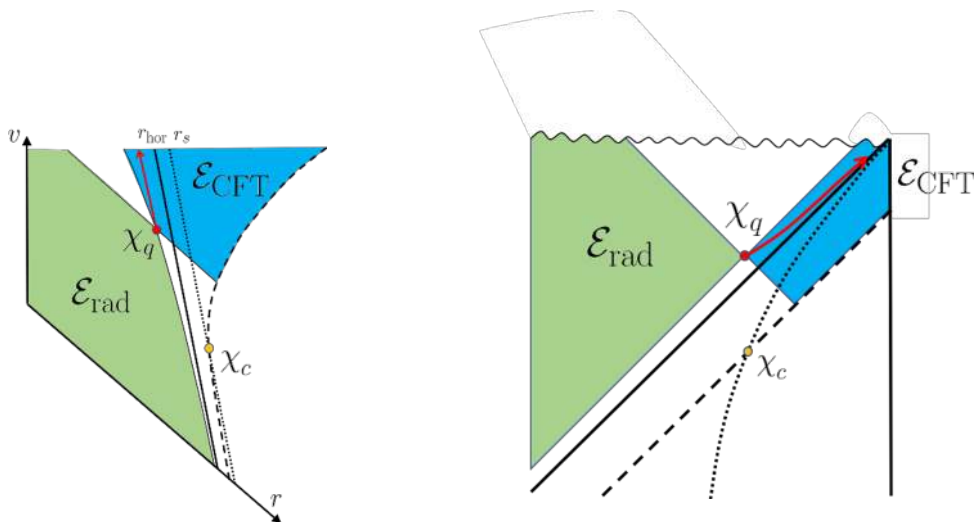
Figure 4. The quantum Ryu-Takayanagi surface χ_q, the classical maximin surface χ_c, and the entanglement wedges $\mathcal{E}_{\text{rad}}$ and $\mathcal{E}_{\text{CFT}}$ of the reservoir and CFT, in Eddington-Finkelstein coordinates (left) and in a Penrose diagram (right). In the interests of simplicity, the Penrose diagram does not include the post-evaporation region, which would be in the top right. The classical maximin surfaces lies at the intersection of the past lightcone (dashed) with the apparent horizon r_s (dotted), which is outside the event horizon. The quantum RT surface, in contrast, lies slightly inside the event horizon. Much of the interior is in the entanglement wedge $\mathcal{E}_{\text{rad}}$ of the reservoir (green), although part of the interior still lies in the entanglement wedge $\mathcal{E}_{\text{CFT}}$ of the CFT (blue). As the black hole continues to evaporate, the RT surface moves forward in infalling time along a spacelike trajectory, following the red arrow. On timescales that are small compared to the evaporation time, it remains a fixed radial distance inside the event horizon.

If we drop terms that are independent of position, it follows from (2.19) that the bulk entropy is given by

$$S_{\text{bulk}} = \frac{c_{\text{evap}}}{6} \log\left(r_{lc}(v) - r\right) - \frac{c_{\text{evap}}\pi v}{6\beta} + \dots \tag{2.26}$$

Because of the rotational symmetry, it is sufficient to show that the surface is extremal under perturbations that preserve the symmetry. Varying the infalling time v, while holding the radius r fixed, we find,

$$0 = \left.\frac{\partial S_{\text{bulk}}}{\partial v}\right|_r + \frac{1}{4G_N}\left.\frac{\partial A}{\partial v}\right|_r, \tag{2.27}$$

$$= \frac{\partial S_{\text{bulk}}}{\partial v}, \tag{2.28}$$

$$= \frac{dr_{lc}/dv}{6(r_{lc} - r)} - \frac{\pi}{6\beta}, \tag{2.29}$$

$$r_{lc} - r = \frac{\beta}{\pi}\frac{dr_{lc}}{dv} \tag{2.30}$$

$$= 2(r_{lc} - r_s). \tag{2.31}$$

In the last equality, we used (2.8). Varying the radius r, at fixed infalling time v, we find,

$$0 = \left.\frac{\partial S_{\text{bulk}}}{\partial r}\right|_v + \frac{1}{4G_N} \left.\frac{\partial A}{\partial r}\right|_v, \tag{2.32}$$

$$= -\frac{c_{\text{evap}}}{6(r_{lc} - r)} + \frac{(d-1)\Omega_{d-1}r_s^{d-2}}{4G_N}, \tag{2.33}$$

$$r_{lc} - r = \frac{2G_N c_{\text{evap}}}{3(d-1)\Omega_{d-1}r_s^{d-2}}, \tag{2.34}$$

$$= \frac{2\beta}{\pi}\frac{dr_s}{dv} = 4(r_s - r_{\text{hor}}), \tag{2.35}$$

where in the last equality we have used (2.17). The quantum extremal surface therefore lies at

$$v = v_c + \frac{\beta}{2\pi}\log 3 = -\frac{\beta}{2\pi}\log\frac{S_{BH}}{c_{\text{evap}}} + O(\beta), \tag{2.36}$$

where v_c is the infalling time of the classical maximin surface found in (2.18), and

$$r = r_s - \frac{\beta}{\pi}\frac{\partial r_s}{\partial v} = r_{\text{hor}} - (r_s - r_{\text{hor}}). \tag{2.37}$$

The extremal surface is twice as far inside the apparent horizon as the event horizon because the entropy of the Hawking radiation produced by the black hole is twice the Bekenstein-Hawking entropy lost by the black hole, which can be easily seen by noting that the energy E and entropy S of black body radiation in two spacetime dimensions are related by

$$E = \frac{1}{2}TS. \tag{2.38}$$

The location of the quantum extremal surface is shown in figure 4, together with the classical maximin surface and the two entanglement wedges.

It is important to note that the entanglement wedge of $\mathcal{H}_{\text{CFT}}$ is bounded by the past lightcone, rather than by the boundary of anti-de Sitter space. This is because the boundary conditions are not deterministic when evolving backwards into the past, without access to $\mathcal{H}_{\text{rad}}$. The region outside the past lightcone is therefore not in the bulk domain of dependence of a spacelike surface connecting the RT surface to the boundary.

The causal wedge is the intersection of the causal past and future of the boundary domain of dependence. Since the boundary time evolution is irreversible, only the future of the boundary is in its domain of dependence. The causal wedge is therefore the intersection of the exterior of the black hole with the future of an infalling lightcone from the boundary. The entanglement wedge contains the causal wedge, as expected.

Of course, we can simply evolve the system back in time using standard reflective boundary conditions; on the boundary, this corresponds to using the ordinary, uncoupled Hamiltonian for the CFT. For this spacetime, both the future and past of the boundary are in its domain of dependence and so the entire exterior of the black hole will be in the causal wedge. One might worry that the entanglement wedge will not contain the causal wedge.

Specifically, in our original spacetime, a signal could easily travel backwards in time from the entanglement wedge of $\mathcal{H}_{\text{rad}}$ to the boundary. If this was still possible under an
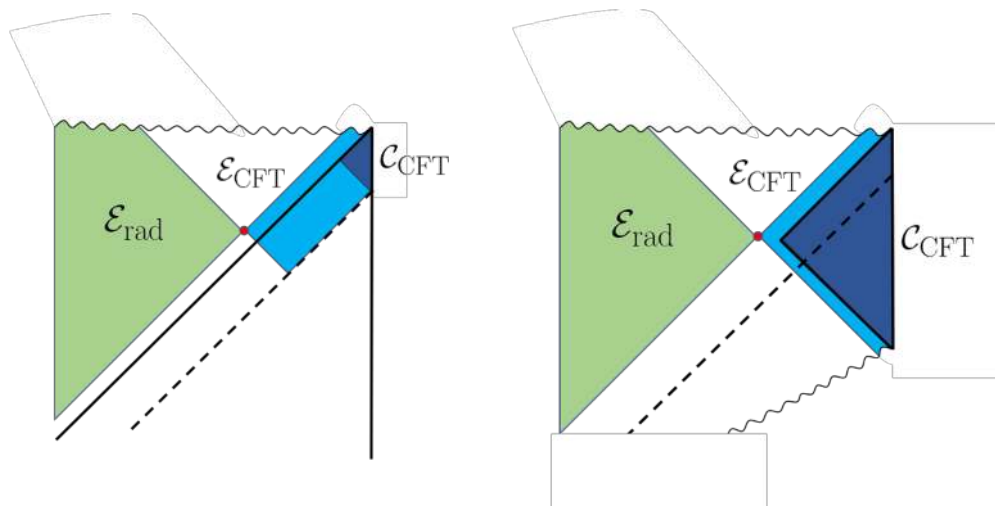
Figure 5. Because the absorbing boundary conditions are only deterministic when evolving forwards in time, the domain of dependence of the boundary is the future boundary. The entanglement wedge $\mathcal{E}_{\mathrm{CFT}}$ of the CFT (light blue) contains the causal wedge $\mathcal{C}_{\mathrm{CFT}}$ (dark blue). When time is reversed using standard reflective boundary conditions, the entanglement wedge still contains the causal wedge because the backreaction on the geometry creates a white hole.

evolution where the two systems $\mathcal{H}_{\mathrm{rad}}$ and $\mathcal{H}_{\mathrm{CFT}}$ were uncoupled, then we would have a serious problem.

However this fails to take into account the backreaction on the spacetime geometry that happens when we change the dynamics.[35] It is easy to see that the geometry of the spacetime must change when we change the boundary conditions. Without the Hawking radiation from $\mathcal{H}_{\mathrm{rad}}$, the black hole cannot grow indefinitely as we evolve the state backwards into the past; it does not have the energy to do so.

Instead, the discontinuity in the outgoing modes will create a shell of high energy density at the past lightcone; the energy of this shell will be proportional to the number of modes c_{evap} that were extracted. As the shell evolves back into the past, it will be blueshifted, creating large backreaction on the geometry once it is a distance $O(c_{\mathrm{evap}} G_N)$ from the black hole. This will create a white hole with Schwarzschild radius $O(c_{\mathrm{evap}} G_N)$ larger than the original black hole. The Ryu-Takayanagi surface will now lie slightly inside the bifurcation horizon of the new spacetime and the entanglement wedge of the CFT will continue to contain the causal wedge.

2.3 Hayden-Preskill and the Page curve

In this subsection, we show how the Ryu-Takayanagi surface, calculated in section 2.2, explains properties of black hole evaporation, such as the Hayden-Preskill decoding criterion and the Page curve, that have been conjectured based on simple toy models of black hole evaporation.

---

[35]See [31] for discussion of essentially the same effect in terms of the dynamics of the boundary particle in $(1 + 1)$-dimensional gravity.
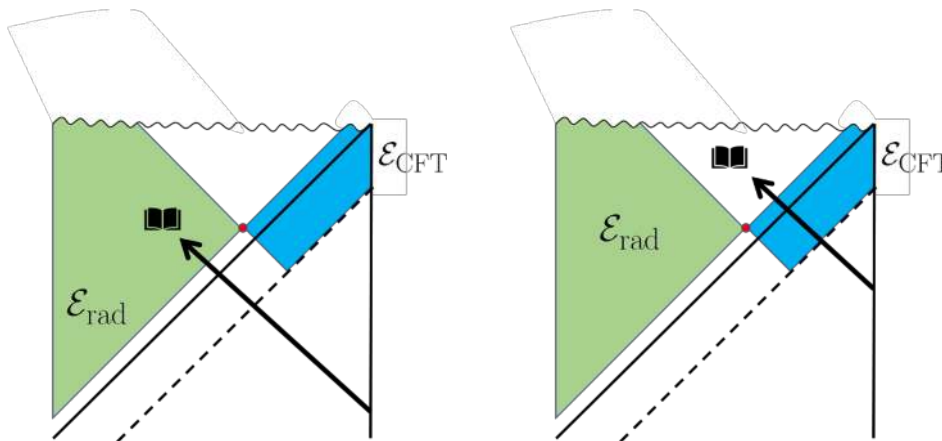
Figure 6. If a diary was thrown into the black hole more than the scrambling time into the past (left), it will now lie in the entanglement wedge of the reservoir $\mathcal{H}_{\text{rad}}$ and can in principle be decoded using only the Hawking radiation. A diary thrown into the black hole more recently (right) remains in the entanglement wedge of, and encoded in, the CFT.

We start with the Hayden-Preskill decoding criterion [8]. This says that, if an unknown, small, light diary is thrown into a black hole, whose state is known, at an early stage in its evaporation, the diary can be decoded from the Hawking radiation almost immediately after the Page time. If the small diary is instead thrown into the black hole *after* the Page time, it can be decoded from the Hawking radiation after waiting for the scrambling time.

This indeed exactly what we see from entanglement wedge reconstruction. After the Page time, the quantum extremal surface lies near the black hole horizon at an infalling time

$$v = -\frac{\beta}{2\pi} \log \frac{S_{BH}}{c_{\text{evap}}} + O(\beta). \tag{2.39}$$

Assuming $c_{\text{evap}} = O(1)$, this is exactly one scrambling time, plus subleading corrections, before the current time. A diary thrown into the black hole before this time lies in the entanglement wedge of, and can be decoded from, the reservoir $\mathcal{H}_{\text{rad}}$ containing the Hawking radiation. Anything thrown in after this time lies in the entanglement wedge of the CFT. This is shown in figure 6.

Of course, if we actually throw a diary into the black hole, it will have non-zero energy and will therefore backreact on the geometry. So long as the energy of the diary is $O(1)$, the backreaction will only change the horizon area by an $O(G_N)$ amount. However, the evaporation of the black hole changes the horizon area by an $O(G_N)$ amount over one thermal time β. Hence it is reasonable to expect that the backreaction will only affect the delay until the diary can be reconstruction from the Hawking radiation by a subleading $O(\beta)$ amount. When we study the reconstruction of large diaries in section 3.4, we will see that this is indeed the case.

In addition to the well-known scrambling time delay in recovering information thrown into a black hole, (2.39) has a small logarithmic correction based on the rate c_{evap} at which Hawking radiation is extracted from the black hole. While we postpone any formal

calculation to future work, it is easy to see heuristically that this is consistent with the boundary dynamics of the theory.

In a fast scrambling system, the number of degrees of freedom that a 'simple' initial perturbation influences grows exponentially with time. This is sometimes described as an 'epidemic' where each 'infected' qubit infects an $O(1)$ number of other qubits in each timestep (which in our case corresponds to an $O(\beta)$ timescale). After the time given in (2.39), a $O(1/c_{\text{evap}})$ fraction of the degrees of freedom will be infected. However, the number of degrees of freedom extracted per timestep is $O(c_{\text{evap}})$. So it is reasonable to expect that an observer with access to the extracted Hawking radiation should be able to detect the perturbation, and hence decode the diary, after the time given in (2.39).

We emphasize that the state of the diary being encoded in the early Hawking radiation does not mean that the diary has been magically extracted out of the interior of the black hole and into the reservoir $\mathcal{H}_{\text{rad}}$. It is 'still' in the interior. There exists a single spacetime that describes the evaporating black hole (which is semiclassical everywhere except for regions of high curvature). In this spacetime, the diary falls into the black hole and keeps falling until it approaches the singularity and the semiclassical spacetime breaks down. It cannot 'no longer' have this worldline — that's not how spacetime works. Spacetime does not change over time; it describes changes over time.

Instead, the encoding of the diary in the Hawking radiation should be understood in terms of the usual story of holography. An object sitting in the middle of the bulk is not 'actually' at the boundary; it is in the middle of the bulk. In the effective field theory that describes the bulk, it is an independent degree of freedom from all the fields at asymptotic infinity.

Nonetheless, by manipulating the fields at asymptotic infinity in a sufficiently complicated way, we can make the bulk effective field theory breakdown and thereby manipulate the object in the middle. Microscopically, the fields at asymptotic infinity contain all the degrees of freedom of the theory.

We should therefore not be too surprised that, at a microscopic level, the diary in the interior is not an independent degree of freedom from the radiation in the reservoir, and hence, with sufficiently complicated manipulations of the reservoir, one can, in principle, manipulate the diary.[36]

We have understood the Hayden-Preskill decoding criterion by analysing the entanglement wedge, $\mathcal{H}_{\text{CFT}}$ or $\mathcal{H}_{\text{rad}}$, that particular *ingoing* modes are in. The Page curve, and a resolution of the firewall paradox, will follow from analysing the entanglement wedge of *outgoing* modes.

We first note that, since we know the location of the Ryu-Takayanagi surface, it is easy to find the entanglement entropy between the Hawking radiation reservoir $\mathcal{H}_{\text{rad}}$ and the conformal field theory $\mathcal{H}_{\text{CFT}}$ (and hence the black hole) using the Ryu-Takayanagi formula. After the Page time, the entanglement entropy is given to leading order by the Bekenstein-Hawking entropy $A_{\text{hor}}/4G_N$ of the black hole, plus a subleading correction

---

[36]This is not to say that there aren't serious conceptual questions that remain to be understood about the relationship between the microscopic boundary theory and the effective bulk theory; it is just that they are fundamentally they same conceptual problems that always exist in holography, even without any black holes.
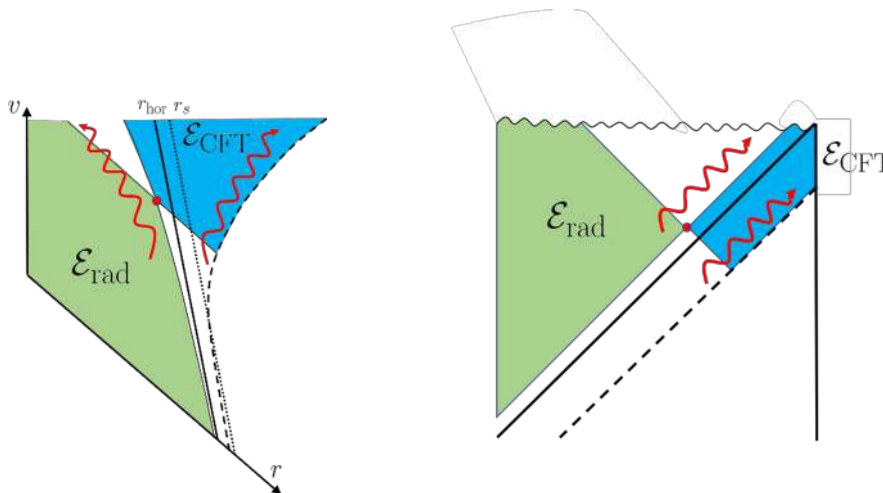
Figure 7. The next Hawking modes to escape the black hole and be extracted into $\mathcal{H}_{\mathrm{rad}}$ will be entangled with interior modes that are mostly in the entanglement wedge of the reservoir $\mathcal{H}_{\mathrm{rad}}$. This causes the entanglement between the black hole and the reservoir to decrease as the new radiation is extracted, in accordance with the Page curve. (Left: Eddington-Finkelstein coordinates; right: a Penrose diagram.)

from the bulk entropy term. Since we have already calculated the entanglement entropy before the Page time, at the start of this section, we have therefore successfully derived the entire Page curve.

However, on its own, this is somewhat unsatisfying. It does not explain *why* moving outgoing Hawking modes, which we naively thought were unentangled with the earlier radiation, from the CFT to the reservoir $\mathcal{H}_{\mathrm{rad}}$ should somehow decrease the entanglement between the two. It does not explain how the AMPS firewall paradox [10] is avoided.

Fortunately, in addition to the Ryu-Takayanagi formula, we also know about entanglement wedge reconstruction. Again, we start with heuristic arguments and then progress to more precise statements. Consider a Hawking mode that escapes the black hole slightly into the future. As shown in figure 7, we can heuristically think of this mode as being entangled with a partner mode behind the horizon. The partner modes will be in the entanglement wedge of $\mathcal{H}_{\mathrm{rad}}$. Hence moving this Hawking quanta from $\mathcal{H}_{\mathrm{CFT}}$ to $\mathcal{H}_{\mathrm{rad}}$ will *decrease* the entanglement between the two. The same ER=EPR resolution [18] to the firewall paradox that worked for the two-sided black hole also works for a one-sided evaporating black hole.

Of course, this is only an approximate heuristic picture. For free fields, Rindler modes, with a given Rindler frequency, outside and inside the horizon are indeed perfectly entangled with one another. However such modes are completely delocalised within the exterior and interior respectively. Localised modes outside the horizon will not be perfectly entangled with their reflection inside the horizon, unless the modes have support in only a very narrow range of Rindler frequencies, and hence are delocalised across a large region in Rindler units.

In this case, since part of the interior is in the entanglement wedge of the CFT, we cannot find modes, with support in only a narrow range of Rindler frequencies, whose reflection inside the horizon will be entirely in the entanglement wedge of $\mathcal{H}_{\mathrm{rad}}$. We should therefore expect that the thermal outgoing radiation will be mostly entangled with the reservoir $\mathcal{H}_{\mathrm{rad}}$,
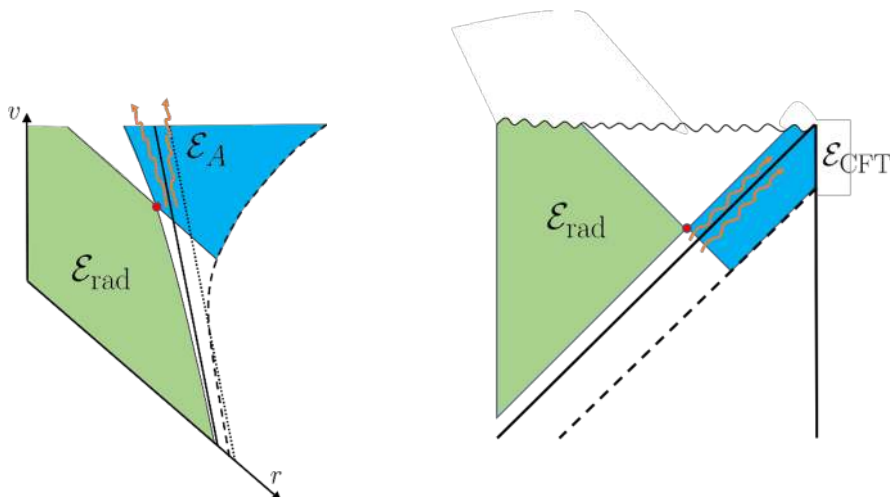
Figure 8. Hawking modes that will escape the black hole only an $O(\beta)$ time into the future are entangled with interior modes that are almost entirely in the entanglement wedge of the CFT. This is very different from a simple random unitary toy model, but is consistent with toy models where the thermodynamic entropy (i.e. number of qubits) increases over time. (Left: Eddington-Finkelstein coordinates; right: Penrose diagram.)

but also be somewhat entangled with the CFT. Moving the Hawking quanta from $\mathcal{H}_{\text{CFT}}$ to $\mathcal{H}_{\text{rad}}$ will decrease the entanglement entropy, but by less than the entropy of the Hawking quanta themselves. This agrees with the Page curve, since the total thermodynamic entropy of the CFT and reservoir is increasing over time and hence

$$\frac{1}{4G_N}\frac{dA_{\text{hor}}}{dv} > -\frac{dS_{\text{rad}}}{dv}, \tag{2.40}$$

even at leading order. We will do a formal calculation that finds perfect agreement between the bulk entanglement structure and the Page curve below.

Interestingly, and somewhat counterintuitively, Hawking quanta that escape at a time only $O(\beta)$ into the future will be almost perfectly entangled with interior modes that lie almost entirely in the entanglement wedge of $\mathcal{H}_{\text{CFT}}$, as shown in figure 8. (In this case, by making the outgoing Hawking mode escape sufficiently far, although still an $O(\beta)$ time, into the future, we can indeed consider outgoing wavepackets with support in only a narrow range of Rindler frequencies, but with 'mirror' interior modes that are entirely within the entanglement wedge of $\mathcal{H}_{\text{CFT}}$.) They will therefore be almost completely *unentangled* with the reservoir $\mathcal{H}_{\text{rad}}$.

This is in sharp contrast with the most naïve random unitary toy models of black hole evaporation. If our model consists of a single random unitary acting on the initial black hole state and then qubits being released one by one as Hawking radiation, we find that any single qubit of Hawking radiation is almost perfectly entangled with *any* set of more than half the qubits. Hence, if we collect Hawking radiation until after the Page time, throw away qubits of Hawking radiation for a while, and then finally collect one more qubit of Hawking radiation, we still find that the additional qubit of Hawking radiation is almost perfectly entangled with the early Hawking radiation that we collected.
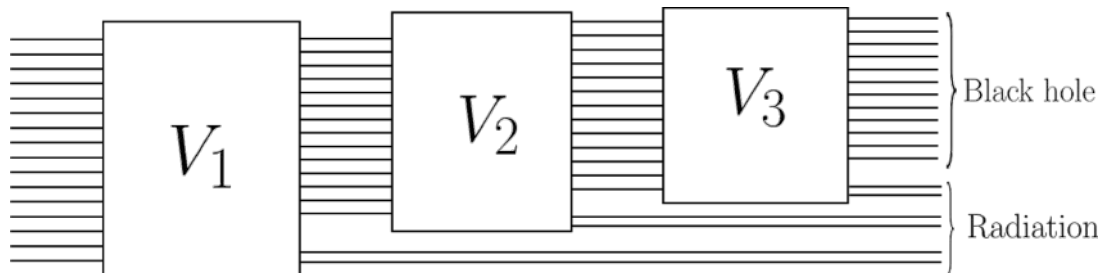
Figure 9. A simple toy model of black hole evaporation that takes into account the increase in thermodynamic entropy as the black holes evaporates. At each timestep, two qubits escape the black hole as Hawking radiation, but then a random isometry is applied to the black hole that increases the number of qubits by one. The total number of qubits (corresponding to the total thermodynamic entropy) therefore increases by one qubit at each timestep. Even if we collect the radiation qubits until long after the Page time, qubits that are radiated only a few timesteps into the future will be completely unentangled with the radiation that we have collected.

However such a model does not take into account the fact that the total combined thermodynamic entropy of the black hole and Hawking radiation, which corresponds in the toy model to the total number of qubits, is strictly increasing over time as the black hole evaporates. A more sophisticated toy model, which does take into account this thermodynamic entropy increase, involves a series of nested random isometries, as shown in figure 9 [38].

At each step, qubits are extracted into the Hawking radiation, but then a random isometry is applied to the black hole so that the number of black hole qubits decreases by less than the number of Hawking radiation qubits increases. It can easily be seen using Page's theorem [60] that an additional qubit of Hawking radiation will be almost totally uncorrelated with the early Hawking radiation, so long as a large, but $O(1)$, number of qubits are thrown away in between.[37]

If the infalling modes are in a thermal state at a temperature equal to the temperature of the black hole, there will be no increase in thermodynamic entropy. We study finite-temperature infalling modes in appendix B, where we indeed find that, if the temperature of the infalling modes is equal to the black hole temperature, the Hawking radiation will be completely unentangled with any CFT degrees of freedom, even when it escapes far into the future. We also find, in several separate cases such as thermal infalling modes and pure infalling modes with constant energy density, that information stops escaping the black hole and the Hawking radiation becomes completely thermal at exactly the moment when this becomes consistent with unitarity.

---

[37]We call the early radiation Hilbert space $\mathcal{H}_E$, and the black hole Hilbert space, when we stop collecting radiation, $\mathcal{H}_{BH}$. There is then a random isometry $V : \mathcal{H}_{BH} \to \mathcal{H}_T \otimes \mathcal{H}_Q \otimes \mathcal{H}_Z$, where $\mathcal{H}_T$ contains the thrown away radiation, $\mathcal{H}_Q$ is the final collected qubit and $\mathcal{H}_Z$ contains the remaining black hole qubits. When we stop collecting the Hawking radiation, the state $|\psi\rangle \in \mathcal{H}_{BH} \otimes \mathcal{H}_E$ will be close to maximally entangled. Since we are after the Page time, $|\mathcal{H}_E| \gg |\mathcal{H}_{BH}|$ and so $|\psi\rangle$ only has support in a subspace $\tilde{\mathcal{H}}_E \subseteq \mathcal{H}_E$ with $|\tilde{\mathcal{H}}_E| = |\mathcal{H}_{BH}|$. However, since $|\tilde{\mathcal{H}}_E \otimes \mathcal{H}_Q| \ll |\mathcal{H}_T \otimes \mathcal{H}_Z|$, the reduced density matrix of the state $V |\psi\rangle$ on $\mathcal{H}_E \otimes \mathcal{H}_Q$ will be very close to maximally mixed. There will be essentially no entanglement, or even correlation, between the early radiation and the final collected qubit.

Having described heuristically how entanglement wedge reconstruction allows us to avoid the firewall paradox, let us now do a more formal calculation of the change in entanglement entropy from extracting bulk Hawking modes. In fact, we shall prove that this change in entanglement entropy will necessarily always agree with the Ryu-Takayanagi formula. However, we first start by calculating it explicitly.

We want to calculate the change in entanglement entropy from outgoing modes, over some small time range δv, being transferred from $\mathcal{H}_{\mathrm{CFT}}$ to $\mathcal{H}_{\mathrm{rad}}$. We need the time δv to be small, because otherwise the Ryu-Takayanagi surface, and hence the entanglement wedges, of $\mathcal{H}_{\mathrm{CFT}}$ and $\mathcal{H}_{\mathrm{rad}}$ will depend on whether the transferred modes are included in $\mathcal{H}_{\mathrm{CFT}}$ or $\mathcal{H}_{\mathrm{rad}}$. By making δv very small, we ensure that all bulk modes (other than the transferred ones) are encoded in one of the two Hilbert spaces, even if the transferred modes themselves are not counted as part of either Hilbert space. One can then calculate the change in entanglement over longer time periods by integrating these infinitesimal changes.

Outgoing modes that are between the Ryu-Takayanagi surface and the past lightcone of the boundary are encoded in the CFT, while all other outgoing modes are encoded in the reservoir $\mathcal{H}_{\mathrm{rad}}$.[38] As discussed at the beginning of this section, since the overall state of the outgoing modes is pure, we will find the same change in entanglement entropy if we look at the change in the entropy of the outgoing modes in $\mathcal{H}_{\mathrm{CFT}}$, or the entropy of the outgoing modes in $\mathcal{H}_{\mathrm{rad}}$. However, since the modes in the CFT consist of a single interval, the change in their entropy is more natural to calculate.

From (2.8), it is easy to see that extracting the Hawking radiation for an additional time δv will move the radius r_{lc} of the outgoing lightcone by

$$\delta r_{lc}(v) = -\frac{2\pi \delta v}{\beta}(r_{lc}(v) - r_{\mathrm{hor}}(v)). \tag{2.41}$$

Assuming we keep the cut-off on the extracted outgoing modes constant, extracting the additional Hawking radiation will change the cut-off ε_2 in units of r by

$$\delta \varepsilon_2 = -\frac{2\pi \delta v}{\beta}\varepsilon_2. \tag{2.42}$$

To derive this equation, we note that the cut-off ε_2 only depends on the difference between the infalling time at which the radiation is extracted and the infalling time of the quantum extremal surface. Hence, extracting radiation for an additional time δv has the same effect on ε_2 as moving the extremal surface backward in infalling time by δv; (2.42) is therefore an immediate consequence of (2.25).

---

[38]We are ignoring the fact that entanglement wedge reconstruction is only approximate here. Since the separation between the outgoing lightcone and the extremal surface grows linearly in the limit of large c_{evap}, the effect of the reconstruction errors should become small in this limit. Furthermore, it is reasonable to hope that the effect of errors in the reconstruction of bulk modes on each side will cancel and so we will still find the correct answer even at small c_{evap}. As we shall see, it seems that this is indeed the case.

Using (2.19), we find that

$$\delta S = \frac{c_{\text{evap}}}{6} \frac{\delta r_{lc}}{r_{lc} - r} - \frac{c_{\text{evap}}}{12} \frac{\delta \varepsilon_2}{\varepsilon_2} \tag{2.43}$$

$$= -\frac{c_{\text{evap}} \pi \, \delta v \, (r_{lc}(v) - r_{\text{hor}}(v))}{3 \, (r_{lc} - r)} + \frac{c_{\text{evap}} \pi \, \delta v}{6} \tag{2.44}$$

$$= -\frac{c_{\text{evap}} \pi \delta v}{12 \beta} = \frac{1}{4 G_N} \frac{dA_{\text{hor}}}{dv} \delta v, \tag{2.45}$$

where in the second equality we used (2.41) and (2.42), in the third equality we used (2.31) and (2.35) and in the last equality we used (2.16). The change in entanglement entropy exactly agrees with the change in entanglement entropy that one finds using the Ryu-Takayanagi formula.

At first glance, these two methods of calculating the change in entanglement entropy appear very different, despite the perfect quantitative agreement between them. In the Ryu-Takayanagi formula, the change comes from the change in the horizon area of the black hole, with its somewhat mysterious association with entropy, while, in (2.43), the change occurs because outgoing bulk modes are entanglemed with other outgoing bulk modes in the entanglement wedge of $\mathcal{H}_{\text{rad}}$.

However, it is not a coincidence that they give the same answer. The Ryu-Takayanagi formula is really calculating the change in the generalised entropy $A/4G_N + S_{\text{bulk}}$ of the Ryu-Takayanagi surface, not just the change in the area; it is just that, in this case, the bulk entropy happens to stay approximately constant (because of the approximate time translation invariance of the evaporation process). The bulk entropy calculation can also be thought of as a change in generalised entropy; just one in which the RT surface for which the generalised entropy is evaluated, and therefore the area term, stay fixed.

In the Ryu-Takayanagi formula calculation, the change in entropy is given by the difference between the new generalised entropy for the new Ryu-Takayanagi surface and the old generalised entropy for the old Ryu-Takayanagi surface. In the bulk entanglement calculation, the change in entropy is the difference between the new generalised entropy and the old generalised entropy, when both are evaluated using the old Ryu-Takayanagi surface. However, by definition, the generalised entropy is constant at leading order if we perturb the Ryu-Takayanagi surface. Since we need δv to be infinitessimally small to do the bulk entanglement calculation, the two calculations will always give the same answer. Because the Ryu-Takayanagi surface is a quantum extremal surface, there can never be a firewall paradox. The bulk entanglement structure will always be consistent with the Ryu-Takayanagi formula.

2.4 Greybody factors

As we discussed in section 2.2, there is nothing genuinely unphysical about evaporating a black hole in AdS/CFT by extracting black-body Hawking radiation from well inside the zone. However, if we eventually want to understand four dimensional black holes in flat space that evaporate naturally (without an external super-observer extracting Hawking

radiation from near the horizon), it is important to understand what happens when there are non-trivial greybody factors.

Although we will not be able to explicitly calculate the location of the quantum extremal surface when greybody factors are present, it turns out that we will still be able to derive both the Hayden-Preskill decoding criterion and the Page curve.[39]

We first recall our argument, from the very beginning of this section, that the maximin prescription implies that the Ryu-Takayanagi surface must become non-empty at the Page time, even when the greybody factors are non-trivial. After this time, assuming that the quantum maximin prescription is valid, there must exist a non-empty quantum extremal surface.[40] We shall show that this is indeed the case

What can we say about the location of this quantum extremal surface? We know that the entanglement wedge must contain the causal wedge [29], so the extremal surface must lie in the interior of the black hole.[41] However we also know that there does not exist a classical extremal surface anywhere in this spacetime. This means that, at the non-empty quantum extremal surface, the gradient of the bulk entropy term must be $O(1/G_N)$ (at least in Eddington-Finkelstein coordinates where the gradient of the area is $O(1)$ everywhere).

Since the bulk entropy itself is $O(1)$, the only way that this can happen is if the extremal surface is very close in Eddington-Finkelstein coordinates to a point where the bulk entropy diverges. The extremal surface must therefore approach the past lightcone of the boundary, which is always outside the event horizon and only approaches the event horizon at infalling times that are far into the past.[42] In Eddington-Finkelstein coordinates, the quantum extremal surface must therefore both approach the event horizon, with respect to the radius r, and diverge into the infinite past, with respect to the infalling time v, in the limit $G_N \to 0$. Again, we shall explicitly verify that this is the case.

It is helpful at this point to switch from Eddington-Finkelstein coordinates to lightlike, Kruskal-Szekeres-like coordinates. At radii close to the event horizon, and over infalling timescales that are small compared to the evaporation time, the metric of the evaporating

---

[39]Since the entanglement entropy of gravitons is not understood, we shall still assume that no graviton modes are extracted using the absorbing boundary conditions and hence that their entanglement entropy can be ignored. Of course, in flat space, gravitons will always contribute to the Hawking radiation, so understanding their entanglement entropy precisely is an important task for future work. However, since the relevant graviton modes simply become ordinary light scalar fields in a $(1+1)$-dimensional reduction of the evaporation, as do all the other bosonic modes that contribute to the evaporation, it seems reasonable to expect that their contributions to the bulk entropy will be qualitatively the same as any other mode.

[40]In fact, since we continue to assume rotational symmetry, this argument would not actually require the full power of the assumption of quantum maximin. Instead, we can restrict our maximisation and minimisation to rotational symmetric Cauchy slices and surfaces χ respectively, thereby avoiding most of the subtleties that would be involved in defining the quantum maximin prescription and showing its equivalence to the quantum extremal surface prescription.

[41]We shall prove explicitly, later in this section, that the quantum extremal surface that we find is indeed inside the event horizon of the black hole.

[42]Technically, the bulk entropy will also diverge near the future lightcone. However there cannot be a quantum extremal surface near the future lightcone, because, at the future lightcone, the bulk entropy will only diverge as function of the infalling time v, while $dA = (d-1) \, r_s^{d-2} \Omega_{d-1} dr$.

black hole in Eddington-Finkelstein coordinates is given by

$$ds^2 = -\frac{4\pi}{\beta}(r - r_s(v))dv^2 + 2dv\,dr + r^2 d\Omega_{d-1}^2, \qquad (2.46)$$

where we can assume that the inverse temperature β and the evaporation rate dr_s/dv are constant at leading order. Substituting the Kruskal-like coordinates,

$$V = \frac{\beta}{2\pi}\exp(2\pi v/\beta), \qquad (2.47)$$

and

$$U = (r_{\mathrm{hor}}(v) - r)\exp(-2\pi v/\beta), \qquad (2.48)$$

where $r_{\mathrm{hor}} = r_s + \beta(dr_s/dv)/2\pi$ as in (2.11), we find

$$ds^2 = -2dU\,dV + r^2(U, V)d\Omega_{d-1}^2. \qquad (2.49)$$

Note that the definition (2.48) is only intended to be valid in the near horizon region where outgoing lightrays escape exponentially in Eddington-Finkelstein coordinates. More generally, the coordinate U should be defined in the exterior region by

$$U \propto -\exp(-2\pi u/\beta), \qquad (2.50)$$

where u is the boundary time at which an outgoing lightray reaches the boundary, and then the metric should be analytically extended to the interior where $U > 0$. This will ensure that the coordinates U, V are exactly lightlike everywhere. However, we are only interested in the near horizon region where the definition (2.48) and the metric (2.49) are valid.[43] Note that $V > 0$ everywhere; the infinite past with respect to infalling time v corresponds to the limit $V \to 0^+$. Also note that the past lightcone of the current boundary time (i.e. $v = 0$) is at $U_{l.c.} = -O(r_s)$.[44]

Our basic strategy will be to show that $\partial S_{\mathrm{bulk}}/\partial U$ should approach a well-defined $O(1)$ limit as $V \to 0$, whereas

$$\frac{1}{4G_N}\frac{\partial A}{\partial U} = O\left(\frac{V}{G_N}\right).$$

We will also find that, in the same limit, $\partial S_{\mathrm{bulk}}/\partial V = O(1/V)$, while

$$\frac{1}{4G_N}\frac{\partial A}{\partial V} = O\left(\frac{1}{G_N}\right) + O\left(\frac{1}{V}\right).$$

We will therefore be able to argue that the quantum extremal surface must be at some fixed U that is independent of G_N and at $V = O(G_N)$, which corresponds to an infalling time exactly one scrambling time (plus subleading corrections) into the past.

---

[43]Even within the near horizon region, our conventions for U and V differ from the more standard conventions for Kruskal coordinates in AdS space by constant factors [54]. However, within the near horizon region, our convention will be somewhat more convenient.

[44]As discussed in section 2.1, for near-extremal black holes, we actually have $U_{l.c} = -O(r_s^2/\beta)$. Also, recall that we are assuming for convenience that the past lightcone escapes the *near-horizon* region at $v = 0$. Hence for parametrically small AdS-schwarzschild black holes, the current boundary time is really $v = \pi l_{\mathrm{AdS}} + O(\beta)$. These subtleties will be unimportant for our purposes; the key point is that $U_{l.c.}$ is independent of G_N.

As with the Eddington-Finkelstein coordinates r, v, the metric (2.49) is approximately constant in the near horizon region in terms of the coordinates U, V. We can therefore consistently cut-off ingoing modes at the quantum extremal surface with a constant cut-off in units of V, and outgoing modes with a constant cut-off in units of U. Since

$$\frac{\partial V}{\partial v} = \exp(2\pi v/\beta), \tag{2.51}$$

and

$$\frac{\partial U}{\partial r} = \exp(-2\pi v/\beta), \tag{2.52}$$

have non-trivial infalling-time dependence, the gradient of the entropy of either the ingoing or outgoing modes *alone* will depend on the set of cut-offs that we use. In particular, in section 2.2, when the cut-offs were constant in Eddington-Finkelstein coordinates, there was an increase in bulk entropy, when moving the RT surface forwards in infalling time along an outgoing lightcone, that came from outgoing bulk modes . With constant cut-offs in Kruskal-like coordinates, the same increase in bulk entropy exists, but it comes from the infalling modes. The gradient of the total bulk entropy is the same in both cases.

An advantage of using constant cut-offs in units of U and V is that outgoing modes are not blueshifted, with respect to U, as we evolve them backwards in time. The outgoing modes that are contained in the entanglement wedge of $\mathcal{H}_{\text{CFT}}$ will be determined only by U, and we don't have to worry about blueshifting the cut-off at the outgoing lightcone. If the cut-off at the quantum extremal surface is also constant in units of U, the entropy of the outgoing modes in the entanglement wedge of the CFT will be independent of V. To calculate $\partial S_{\text{bulk}}/\partial V$ we therefore only need to worry about the ingoing modes near the quantum extremal surface.

In section 2.2, the infalling modes were in the vacuum state with respect to the infalling time v. We therefore argued that, so long as the cut-off was constant in units of v, the gradient of the entropy of the infalling modes would tend to zero as the quantum extremal surface diverged into the infinite past in the semiclassical limit.

When there are non-trivial greybody factors, with part of the Hawking radiation being reflected back into the black hole, the infalling modes will instead be in a *mixed* state that is invariant with respect to translations in infalling time. The mixed-state infalling modes will be purified by outgoing modes that escaped the black hole into the reservoir, as well as modes deep in the interior of the black hole. As shown in figure 10, these modes are all in the entanglement wedge of $\mathcal{H}_{\text{rad}}$.

In the semiclassical limit, when the extremal surface diverges into the infinite past, there will be no entanglement between ingoing modes near the quantum extremal surface and outgoing modes in the entanglement wedge of the CFT. We therefore find

$$\frac{2\pi V}{\beta} \left.\frac{\partial S_{\text{bulk}}}{\partial V}\right|_U = \left.\frac{\partial S_{\text{bulk}}}{\partial v}\right|_U = \frac{c_{\text{evap}}\pi}{6\beta} - \frac{dS_{\text{in}}}{dv}, \tag{2.53}$$

where $dS_{\text{in}}/dv \geq 0$ is the constant entropy per unit infalling time of the infalling modes assuming a constant cut-off in units of the infalling time v. dS_{in}/dv can in principle be calculated from a numerical approximation of the reflection coefficients for modes escaping
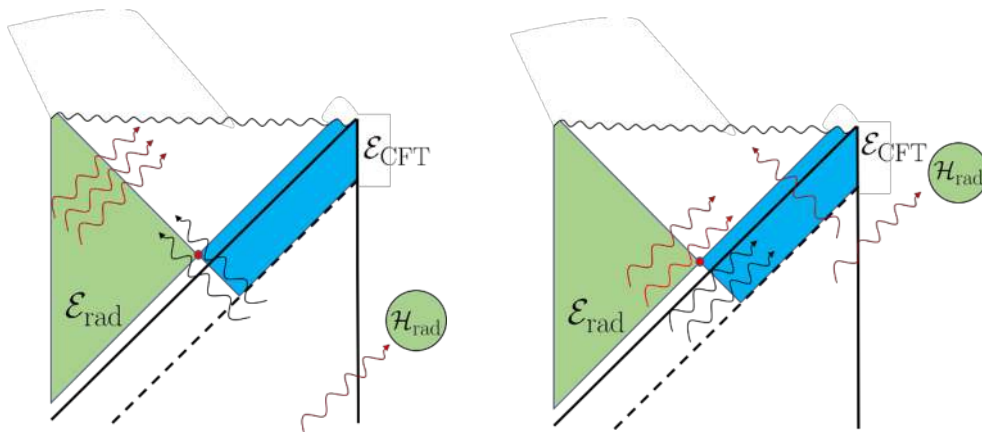
Figure 10. In the left figure, infalling modes near the quantum extremal surface are in a infalling-time-translation invariant mixed state. These mixed state modes (black) are purified by modes deep in the interior, together with modes that escaped into the reservoir (both red). Both of these are in the entanglement wedge of the reservoir $\mathcal{H}_{\mathrm{rad}}$. In the right figure, outgoing modes near the quantum extremal surface (black) are entangled with outgoing modes in the interior, but also with outgoing modes that recently escaped into the reservoir and late-time infalling modes that were reflected back into the black hole (all in red). The first two are in the entanglement wedge of $\mathcal{H}_{\mathrm{rad}}$, while the last is in the entanglement wedge of the CFT. The bulk entropy of the modes in the entanglement wedge of the CFT depends on U no longer depends on U in the same simple way that it did when no greybody factors were present in section 2.2.

the near-horizon zone. (See, for example, similar calculations in [61].) The additional term $c_{\mathrm{evap}}\pi/6\beta$ shows up because the cut-off at the extremal surface should be constant in units of V, rather than constant in units of v, if we want to ignore the outgoing modes. Since

$$\left.\frac{\partial}{\partial V}\right|_U = e^{-2\pi v/\beta}\left.\frac{\partial}{\partial v}\right|_U, \tag{2.54}$$

and the logarithmic divergence $(c_{\mathrm{evap}}/12)\log(1/\varepsilon)$ with respect to the cut-off length ε is universal,[45] we can immediately obtain (2.53).

Formally, there are infinitely many angular momentum modes and so c_{evap} is infinite. However, modes with large angular momentum are almost entirely reflected back into the black hole. The ingoing modes are in a thermal state at the same inverse temperature β as the black hole. They therefore satisfy [57, 58]

$$\frac{dS_{\mathrm{in}}}{dv} = \frac{c_{\mathrm{evap}}\pi}{6\beta}. \tag{2.55}$$

It follows that only the finite number of low angular momentum modes, which actually partially escape the black hole, contribute to (2.53). So long as we include the same modes in calculating dS_{in}/dv that we use in calculating c_{evap}, (2.53) should be independent of the choice of any sufficient large angular momentum cut-off on the modes we consider.

---

[45]Recall that ε is the cut-off on only the ingoing modes, at only one end of an interval.

From (2.47) and (2.48), we have

$$r = r_{\text{hor}}(v) - \frac{2\pi}{\beta} UV, \tag{2.56}$$

and hence

$$\left.\frac{\partial r}{\partial V}\right|_U = \frac{\beta}{2\pi V}\frac{dr_s}{dv} - \frac{2\pi U}{\beta}. \tag{2.57}$$

It follows that

$$\frac{1}{4G_N}\left.\frac{\partial A}{\partial V}\right|_U = \frac{(d-1)r_s^{d-2}\Omega_{d-1}}{4G_N}\left.\frac{\partial r}{\partial V}\right|_U, \tag{2.58}$$

$$= \frac{\beta^2}{2\pi V}\frac{dM}{dv} - \frac{\pi(d-1)\,r_s^{d-2}\,\Omega_{d-1}\,U}{2\beta G_N}. \tag{2.59}$$

In the second equality we used (2.57) and the first law of black hole thermodynamics $\beta dM = dA_{\text{hor}}/4G_N$. The quantum extremal surface must therefore satisfy

$$0 = \frac{1}{4G_N}\left.\frac{\partial A}{\partial V}\right|_U + \left.\frac{\partial S_{\text{bulk}}}{\partial V}\right|_U, \tag{2.60}$$

$$r_{\text{hor}} - r = \frac{2\pi}{\beta} UV = \frac{2G_N\,\beta}{\pi\,(d-1)\,r_s^{d-2}\,\Omega_{d-1}}\left[\beta\frac{dM}{dv} + \frac{c_{\text{evap}}\pi}{6\beta} - \frac{dS_{\text{in}}}{dv}\right], \tag{2.61}$$

where we used (2.53) and (2.59).

It can be shown, as follows, that the right hand side of (2.61) must be non-negative, and so the extremal surface is inside the event horizon of the black hole. Suppose that the infalling modes were in a thermal state at inverse temperature β'. We would then have

$$\frac{dM}{dv} = \frac{dM}{dv} = \frac{c_{\text{evap}}\pi}{12}\Big(\frac{1}{\beta'^2} - \frac{1}{\beta^2}\Big), \tag{2.62}$$

and [57, 58]

$$\frac{dS_{\text{in}}}{dv} = \frac{c_{\text{evap}}\pi}{6\beta'}. \tag{2.63}$$

The right hand side of (2.61) would then be given by

$$\frac{G_N\,\beta^2\,c_{\text{evap}}}{6\,(d-1)\,r_s^{d-2}\,\Omega_{d-1}}\left(\frac{1}{\beta} - \frac{1}{\beta'}\right)^2 \geq 0, \tag{2.64}$$

which is non-negative at any inverse temperature β'.[46] Since thermal states have maximal entropy for any fixed energy flux, the right hand side of (2.61) must therefore be non-negative for any state of the infalling modes, thermal or otherwise. The quantum extremal surface is always inside the event horizon.

So far we have only demanded that the surface be extremal if we vary V at constant U. The quantum extremal surface should also be extremal when we vary U at constant V. From (2.56), we have

$$\frac{1}{4G_N}\left.\frac{\partial A}{\partial U}\right|_V = -\frac{(d-1)\,r_s^{d-2}\,\Omega_{d-1}\,V}{4G_N}. \tag{2.65}$$

---

[46]For detailed calculations of quantum extremal surfaces for finite temperature infalling modes, see appendix B.

What about the variation of the bulk entropy term? By varying U, we change the outgoing modes that are included in the entanglement wedge of the CFT. These outgoing modes are entangled with the other outgoing modes, both inside the quantum extremal surface and outside the past lightcone of the boundary.

In section 2.2, all the outgoing modes that were not in the entanglement wedge of the CFT were in the entanglement wedge of the reservoir. We therefore found

$$\frac{\partial S_{\text{bulk}}}{\partial U}\bigg|_V = \frac{c_{\text{evap}}}{6(U - U_{l.c})}, \tag{2.66}$$

where $U_{l.c.} < 0$ labels the past lightcone of the boundary.[47]

When greybody factors are present, this will no longer be the case, as shown in figure 10. Outgoing modes outside the past lightcone will be partially reflected back into the black hole, and end up as ingoing modes, which *are* in the entanglement wedge of the CFT. The functional form of $\partial S_{\text{bulk}}/\partial U$ will therefore be much more complicated.

However, in the semiclassical limit where the extremal surface diverges into the infinite past, there will still be no entanglement between outgoing modes in the entanglement wedge of the CFT and ingoing modes *near the extremal surface*. The gradient $\partial S_{\text{bulk}}/\partial U$ will therefore be independent of V, so long as V is sufficiently small. Indeed, this follows directly from the fact that $\partial S_{\text{bulk}}/\partial V$ is independent of U, by the symmetry of mixed partial derivatives. We conclude that $\partial S_{\text{bulk}}/\partial U$ has a well-defined limit as $V \to 0$, which is some (presumably complicated) function of U.

If the surface is extremal, we must have

$$0 = \frac{1}{4G_N}\frac{\partial A}{\partial U}\bigg|_V + \frac{\partial S_{\text{bulk}}}{\partial U}\bigg|_V, \tag{2.67}$$

$$\frac{\partial S_{\text{bulk}}}{\partial U}\bigg|_V = \frac{(d-1)\,r_s^{d-2}\,\Omega_{d-1}\,V}{4G_N}, \tag{2.68}$$

$$U\,\frac{\partial S_{\text{bulk}}}{\partial U}\bigg|_V = \frac{\beta}{2\pi}\left[\beta\frac{dM}{dv} + \frac{c_{\text{evap}}\pi}{6\beta} - \frac{dS_{\text{in}}}{dv}\right], \tag{2.69}$$

where in the first equality we used (2.65) and in the last equality we used (2.61). The right hand side of (2.69) is constant over timescales that are small compared to the evaporation time, while the left hand side is a function of U.

If a non-empty quantum extremal surface is to exist in the near horizon region for any sufficiently small G_N, as we expect from the maximin prescription, there must exist a solution U_0 to (2.69). Importantly, since (2.69) does not depend on G_N, this solution must be independent of G_N.

In the simple example from section 2.2 where there are no greybody factors, dM/dv is given in (2.15), $dS_{\text{in}}/dv = 0$ and $\partial S_{\text{bulk}}/\partial U$ is given in (2.66). Hence

$$U_0 = -U_{l.c.}/3, \tag{2.70}$$

in agreement with our calculations in Eddington-Finkelstein coordinates.

---

[47]Of course, our actual calculations in section 2.2 were in Eddington-Finkelstein coordinates, but were equivalent to (2.66). With constant cut-offs in Kruskal-like coordinates, the outgoing bulk entropy $S_{\text{out}} = (c_{\text{evap}}/6)\log((U - U_{l.c})/\sqrt{\varepsilon_1\varepsilon_2})$ where the cut-offs ε_1 and ε_2 are both constant in units of U.

What about when greybody factors are present? In this case, it turns out that we can still calculate $\partial S_{\text{bulk}}/\partial U$ in the limit $U \to \infty$, and this is sufficient to argue that a solution U_0 must exist. For this argument it is simplest to use the Rindler-like coordinates, $u_L = -(\beta/2\pi)\log(U/r_s)$ for $U > 0$ and $u_R = -(\beta/2\pi)\log(|U|/r_s)$ for $U < 0$.[48] Up to a G_N-independent shift, u_R is the time at which an outgoing lightray would reach the boundary, while u_L is its 'mirror' coordinate inside the event horizon.

The outgoing modes are near the horizon are in the thermofield double state with respect to these Rindler coordinates. A key point will be that the thermofield double state only has significant correlation between the left and right region when $u_R = u_L \pm O(\beta)$. If the RT surface is at (U^{RT}, V^{RT}), the interior outgoing modes are in the entanglement wedge for $u_L > u_L^{RT} = -(\beta/2\pi)\log(U^{RT}/r_s)$.

What about the exterior outgoing modes? For $u_R/\beta \gg 0$, they are completely in the entanglement wedge, since they haven't had time to escape. For $u_R/\beta \ll 0$, the situation is more complicated. In this case, the outgoing near-horizon modes are encoded in a combination of the Hawking radiation that has escaped into $\mathcal{H}_{\text{rad}}$, and the reflected ingoing modes at an infalling time $v = u_R + O(\beta)$. For $u_R \gg v^{RT} = (\beta/2\pi)\log(V^{RT}/r_s)$, the reflected modes are in the entanglement wedge, but the escaped Hawking radiation is not. For $u_R \ll v^{RT}$, neither is in the entanglement wedge. Finally, since we are assuming that (2.61) holds, we note that $u_L^{RT} \gg v^{RT}$.

In summary, for $u_L, u_R \ll v^{RT}$, none of the interior or exterior outgoing modes are in the entanglement wedge. For $u_L^{RT} \gg u_L, u_R \gg v^{RT}$, the interior modes are not in the entanglement wedge and only the reflected part of the exterior modes is in the entanglement wedge. For $0 \gg u_L, u_R \gg u_L^{RT}$, the interior modes and the reflected part of the exterior modes are in the entanglement wedge. Finally for $u_L, u_R \gg 0$, all the modes are in the entanglement wedge.

Since the thermofield double state has a local entanglement structure, as discussed above, then we can ignore any effects that come from the finite size of each of these regions, in the semiclassical limit $G_N \to 0$. Instead there are only three contributions to the gradient of the bulk entropy as a function of u_L^{RT}. The first is that increasing u_L^{RT} increases the range of u_L, u_R for which only the reflected part of the interior modes is in the entanglement wedge. This gives a contribution equal to the entropy density dS_{in}/dv of the reflected modes. The second is that increasing u_L^{RT} *decreases* the range of u_L, u_R for which both the interior modes and the reflected part of the exterior modes are in the wedge. Since the tripartite state of interior modes, reflected exterior modes and escaped exterior modes is pure, this gives a contribution equal to $-dS_{\text{rad}}/dv$. Finally, we need the cut-off to be constant in units of U, we means the cut-off length must exponentially grow as a function of u_L. This gives a contribution to the gradient of $-\pi c_{\text{evap}}/6\beta$. We therefore find that

$$U\frac{\partial S_{\text{bulk}}}{\partial U} = -\frac{\beta}{2\pi}\frac{\partial S_{\text{bulk}}}{\partial u_L} = \frac{\beta}{2\pi}\left[\frac{dS_{\text{rad}}}{dv} - \frac{dS_{\text{in}}}{dv} + \frac{\pi c_{\text{evap}}}{6\beta}\right]. \tag{2.71}$$

---

[48]The factors of r_s are included only to make the logarithms dimensionless.

To build some intuition for this formula, note that, in the limit $U \to \infty$ and radius r given in (2.61),

$$\frac{\partial S_{\text{bulk}}}{\partial v}\bigg|_r = -\frac{2\pi U}{\beta} \frac{\partial S_{\text{bulk}}}{\partial U}\bigg|_V + \frac{2\pi V}{\beta} \frac{\partial S_{\text{bulk}}}{\partial V}\bigg|_U = -\frac{dS_{\text{rad}}}{dv}. \tag{2.72}$$

This is in exact agreement with our heuristic picture from way back in figure 2 of Bell pairs, entangled between the interior and the Hawking radiation, that are evenly distributed along the wormhole.

So long as the black hole is evaporating and hence $dM/dv < 0$, it follows from (2.71) that the left hand side of (2.69) should be larger at large U than the right hand side, which is independent of U.[49] Meanwhile, at the horizon, the left hand side is zero, while the right hand side is positive, as shown in (2.64). Hence, assuming that the derivative $\partial S_{\text{bulk}}/\partial U$ is a continuous function of the location of the rotationally symmetric surface, then by the intermediate value theorem there must indeed exist a solution U_0 to (2.69), as we expected from the maximin arguments.[50]

Since we know a solution U_0 must exist, we can substitute it into (2.61) and find that

$$V = \frac{G_N \beta^2}{\pi^2 U_0 (d-1) r_s^{d-2} \Omega_{d-1}} \left[\beta \frac{dM}{dv} + \frac{c_{\text{evap}} \pi}{6\beta} - \frac{dS_{\text{in}}}{dv} \right]. \tag{2.73}$$

Hence

$$v = \frac{\beta}{2\pi} \log \frac{2\pi V}{\beta} = \frac{\beta}{2\pi} \log S_{BH} + O(\beta). \tag{2.74}$$

In the last equality, we assumed for simplicity that all other scales are held fixed in the semiclassical limit $G_N \to 0$. We can therefore derive the Hayden-Preskill decoding criterion even when the greybody factors are non-trivial.

We can also find the Page curve (up to subleading corrections) using the Ryu-Takayanagi formula. Furthermore, our argument from section 2.3, showing that the outgoing modes automactically have exactly the right entanglement to reproduce the Page curve, without any firewall paradox, only used entanglement wedge reconstruction and the fact that the Ryu-Takayanagi surface is an extremum of the generalised entropy. All our main results can therefore be derived, even in the presence of greybody factors.

---

[49]If $dM/dv > 0$, then this will stop being true, and the quantum extremal surface will stop existing, at exactly the moment when the rate of change of the Bekenstein-Hawking entropy $1/4G_N dA_{\text{hor}}/dv = \beta dM/dv$ becomes greater than the rate of increase in the entropy of the Hawking radiation. This is exactly the moment when boundary unitarity becomes consistent with no information ever escaping the black hole. See appendix B for a more detailed discussion of this.

[50]One might wonder whether there could exist multiple solutions and hence multiple non-empty quantum extremal surfaces. We first note that, even if there did exist multiple solutions, the solution that minimised the generalised entropy would be independent of G_N, for sufficiently small G_N, and so we could simple ignore the other solutions. However, in practice, it seems that the left hand side of (2.69) should be a monotonically increasing function of U and so only one solution will exist. If no exterior modes were in the entanglement wedge of the CFT, we would have $\partial S_{\text{bulk}}/\partial U \propto 1/U$ and the left hand side of (2.69) would be constant. The existence of exterior outgoing modes in the entanglement wedge of the CFT should only slow the rate of decay of $\partial S_{\text{bulk}}/\partial U$ as a function of increasing U.

3 State dependence

In deriving the results of section 2, we were careful to focus on a single spacetime where a black hole was formed by collapse in some particular initial state of the matter fields. Although the details of how the black hole was formed did not affect any of our calculations, we always implicitly assumed that those details were known by the observer reconstructing interior operators. We never considered the task of reconstructing a diary from the Hawking radiation with the initial state of the black hole partially, or completely, unknown.

We therefore avoided the issue of whether, and to what extent, reconstructions of bulk interior operators necessarily depend on the initial state of the black hole. Such questions will be the focus of this section. In particular, we will find that this state dependence is crucial in resolving the information paradox. If the interior partners of the late-time Hawking modes were encoded in the early radiation in a state-independent way, there would be no way for the final state of the Hawking radiation to depend on the initial state of the matter falling into the black hole, even if the final state was pure rather than thermal. The state dependence allows the entanglement of early and late radiation to depend on the state, despite the entanglement of late radiation and interior modes being fixed. This allows information to escape.

We begin the section by briefly reviewing results from [38] that show how state dependence can arise in entanglement wedge reconstruction.

3.1 State dependence in entanglement wedge reconstruction

Entanglement wedge reconstruction is best understood in the language of holographic quantum error correction [25]. Bulk operators in AdS/CFT are only well defined within the "code subspace" $\mathcal{H}_{\text{code}} \subseteq \mathcal{H}_{\text{CFT}}$ of boundary states with the correct smooth bulk geometry. The claim of entanglement wedge reconstruction can then be phrased as follows: the noisy quantum channel, mapping states ρ in the code space to their restriction ρ_B to some boundary region B, forms an approximate operator algebra quantum error correcting code for the bulk operators in the entanglement wedge b of B. This means that there exists a decoding channel $\mathcal{D}$ such that, for all states ρ in the code subspace,

$$\mathcal{D}(\rho_B) \approx \rho_b, \tag{3.1}$$

where ρ_b is the restriction of the bulk state ρ to the algebra of operators associated with the entanglement wedge b of B.[51] We can then use the adjoint channel $\mathcal{D}^\dagger$ to map bulk operators ϕ_b, acting within the entanglement wedge, to boundary 'reconstructions' $\phi_B = \mathcal{D}^\dagger(\phi_b)$, acting only on region B, whose action is (approximately) the same as the bulk operator ϕ_b on states in the codespace. It is important to note that the decoding channel $\mathcal{D}$ is in general highly non-unique; two very different boundary reconstructions of the same bulk operator may be both be valid, so long as their action on states in the code subspace is (approximately) the same.

---

[51]Here, restriction can be thought of as a partial trace, although it is really the projection of the state onto the von Neumann subalgebra associated with the bulk region b.

Because the spacetime geometry is dynamical, the entanglement wedge of a given boundary region depends on the bulk geometry of the particular CFT state that we are interested in. Even if we only consider a code space of bulk states with a fixed spacetime geometry, the quantum extremal surface may be state-dependent because of the bulk entropy term. When we say that operators in the entanglement wedge can be reconstructed in the boundary region, we should be careful to specify *which* states need to have their entanglement wedge contain the bulk operator.

The initial derivation of entanglement wedge reconstruction in [23] suggested that one could always reconstruct a bulk operator so long as it was contained within the entanglement wedge of the boundary region for all pure states in the code space of states for which the reconstruction was meant to work. However, this derivation ignored the fact that entanglement wedge reconstruction is only approximate at finite G_N. (More specifically it ignored the fact that the equality between bulk and boundary relative entropies [22] is only approximate at finite G_N.) It should therefore only be trusted for code spaces whose dimension is relatively small (and, in particular, independent of G_N).

A more rigorous derivation of entanglement wedge reconstruction, using the tools of approximate quantum error correction, makes it clear that the bulk operator needs to be contained in the entanglement wedge even for *mixed states* with support only in the code space [24].[52] Rather than directly considering mixed states, it is often more natural, and is mathematically equivalent, to assume that the mixed states are purified by an arbitrary 'reference system' $\mathcal{H}_R$, whose dimension is equal to the code space dimension.[53] In this picture, the bulk operator must be contained in the entanglement wedge of B for all pure states, including entangled states, in $\mathcal{H}_{\text{code}} \otimes \mathcal{H}_R$.

If, instead, the bulk operator is *only* contained in the entanglement wedge of B for all pure states in the code space (with no reference system), entanglement wedge reconstruction will still be possible, but only if we allow the reconstruction to be state-dependent [38].

It is a general fact about quantum error correction that, if exact state-dependent reconstruction of operators is possible all states in a finite-dimensional code space, then exact state-independent reconstruction is also possible for that code space [38, 62].[54] This is why it was sufficient to only consider pure states in [23], where the reconstruction errors were ignored.

Even if the state-dependent reconstruction is only approximate, state-independent reconstruction will necessarily also be possible (with a somewhat larger error), so long as the code space is not too large. In holography, this corresponds to the fact that entanglement with a reference system cannot affect the location of the Ryu-Takayanagi surface in the semiclassical limit, so long as the dimension of the code space is $O(1)$, because any entanglement with the reference system will give a subleading correction to the generalised entropy.

---

[52]For an alternative derivation that is more directly equivalent to the derivation in [23], but which reaches the same conclusion as [24], see [38].

[53]The reference system $\mathcal{H}_R$ should not be confused with the Hawking radiation reservoir $\mathcal{H}_{\text{rad}}$ that we use when studying evaporating black holes. The second is an actual physical system, whereas the first is purely a mathematical 'accounting trick'.

[54]In the Schrödinger picture, which is usually used to describe quantum error correction, this corresponds to the fact that exact universal subspace quantum error correction implies full quantum error correction [62].
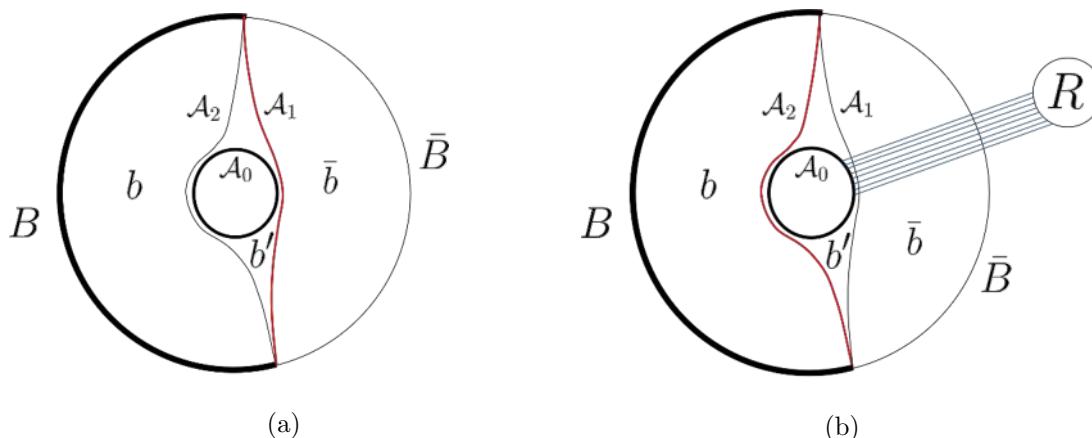
Figure 11. The exterior geometry of a black hole, with horizon area $\mathcal{A}_0$, in AdS/CFT. The boundary is divided into two regions B and $\bar{B}$; there exists a locally minimal surface separating B from $\bar{B}$ on either side of the black hole, with areas $\mathcal{A}_2$ and $\mathcal{A}_1$. These divide the bulk into three regions b, b' and $\bar{b}$, where region b' lies between the two minimal surfaces and contains the black hole. If $\mathcal{A}_2 > \mathcal{A}_1$, region b' is contained in the entanglement wedge of region B for all pure microstates, shown in figure 11a. However, if the black hole is entangled with a reference system with $S_{\rm bulk}/4G_N > \mathcal{A}_2 - \mathcal{A}_1$, as shown in figure 11b, the Ryu-Takayanagi surface will jump to $\mathcal{A}_2$ and the entanglement wedge of B will no longer contain region b'. As a result, a state-independent reconstruction of region b' on region B only exists for code spaces of microstates with dimension $|\mathcal{H}_{\rm code}| \leq e^{(\mathcal{A}_2 - \mathcal{A}_1)/4G_N}$.

In contrast, if the dimension of the code space is very large, for example when one considers a large number of black hole microstates, approximate state-dependent reconstruction may be possible, even when state-independent reconstruction is not.

A simple example of this was studied in [38]. It shows up when one considers code spaces consisting of a large number of black hole microstates, plus bulk degrees of freedom outside the horizon, as illustrated in figure 11. We emphasize that, unlike in the rest of this paper, we will not be interested in the details of the interior of these black hole microstates.

Suppose that we try to reconstruct bulk operators in a simply connected region B consisting of slightly more than half of the boundary. For any pure black hole microstate, the Ryu-Takayanagi surface, with area $\mathcal{A}_1$, lies between the black hole and the complementary region $\bar{B}$. The entanglement wedge of region B contains the black hole.

However, for a two-side black hole, such as the thermofield double state, the homology constraint means that the Ryu-Takayanagi surface lies between region B and the black hole, so long as the area $\mathcal{A}_2$ of this surface is less than the area $\mathcal{A}_1$ plus the horizon area $\mathcal{A}_0$. The entanglement wedge will no longer contain the bulk region b' that lies between the two extremal surfaces.

Replacing the second CFT by an arbitrary reference system cannot affect whether bulk operators in region b' can be reconstructed in region B. Region b' therefore cannot be encoded in region B for any purification of the thermal density matrix. So long as we correctly use the quantum extremal surface prescription to define the Ryu-Takayanagi surface, we do indeed find that this is the case. Instead of classical area and a homology constraint, we now have a large amount of bulk entanglement between the black hole and

the reference system. However, the quantum Ryu-Takayanagi surface, and its generalised entropy, are unchanged.

The black hole does not need to be maximally entangled with the reference system in order to exclude region b' from the entanglement wedge. We only need the entanglement entropy S to satisfy

$$S > \frac{\mathcal{A}_2 - \mathcal{A}_1}{4G_N}. \tag{3.2}$$

At face value, we now have something of a paradox. The bulk region b' is encoded in region B for *any* pure microstate, but for sufficiently entangled state it can be reconstructed on the combination of region $\bar{B}$ and the reference system. By linearity, a reconstruction that works for any pure state will also work for entangled states.[55] But the no cloning theorem says that quantum information can't be simultaneously encoded in region B, and in the combination of region $\bar{B}$ and the reference system.

The resolution, of course, is the fact that entanglement wedge reconstruction can only be made state independent, if the bulk operator is contained in the entanglement wedge even for mixed states with support only in the code space. In this case, the reconstruction will have to be state dependent, precisely when the entropy S of the code space satisfies (3.2). There is therefore no single reconstruction that could be used for the entangled state.

Again, we emphasize that this resolution is only consistent because of the approximate nature of entanglement wedge reconstruction. Otherwise it would be impossible for a state-dependent reconstruction to exist for every state in the code space, without state-independent reconstruction also being possible. There would be no way to evade the paradox described above.

If we make the code space $\mathcal{H}_{\text{code}}$ of microstates as large as possible, so that

$$\log |\mathcal{H}_{\text{code}}| \approx S_{BH} = \frac{\mathcal{A}_0}{4G_N}, \tag{3.3}$$

at leading order, a single reconstruction will only exist for any *subspace* of the code space with dimension less than $|\mathcal{H}_{\text{code}}|^\alpha$, where

$$\alpha = \frac{\mathcal{A}_2 - \mathcal{A}_1}{\mathcal{A}_0}. \tag{3.4}$$

This is an example of something known as an 'α-bit code' [62]. Indeed, many of the results about state-dependence in evaporating black holes that we will derive in this section can be rephrased in the language of [62] as statements about the existence (or non-existence) of α-bit codes for various values of α. However, since the terminology of 'α-bits' was developed for asymptotic quantum resource equalities and is, perhaps, more misleading than clarifying in the present context, we will not use it any further in this paper.

Using the results of [62], it is possible to put strict lower bounds on the size of the non-perturbative error that must exist to avoid a cloning paradox. Specifically we find that

---

[55]For approximate reconstructions, this fact is somewhat non-trivial to prove. However, it is indeed true, up to a dimension-independent increase in the error size [63].

the error in the reconstruction of region b' on region B for a code space $\mathcal{H}_{\text{code}}$ must be at least $e^{-O(\Delta S)}$, where

$$\Delta S = \mathcal{A}_2/4G_N - \mathcal{A}_1/4G_N - \log|\mathcal{H}_{\text{code}}|, \qquad (3.5)$$

is the minimum difference, for any state in the code space, between the generalised entropies of the two extremal surfaces [38].

3.2 State dependence in evaporating black holes

Exactly the same effects that were described in [38] also happen in evaporating black holes. The boundary $\mathcal{H}_{\text{CFT}}$ and the reservoir $\mathcal{H}_{\text{rad}}$ play the roles of the two boundary regions $\mathcal{H}_B$ and $\mathcal{H}_{\bar{B}}$.

Suppose, as in section 2.3, that we throw a small diary into a black hole. This time, however, rather than knowing the initial state of the black hole exactly, we only know that the initial state was in some large code space of possible states.

For example, we can imagine starting with a small black hole, with Bekenstein-Hawking entropy S_{code}, in a *completely unknown* state. If we then throw a large amount of additional energy (this time in a known state) into this black hole, we end up with a larger black hole, whose state is partially unknown. It lies in a code space of possible states, which has entropy $\log|\mathcal{H}_{\text{code}}| = S_{\text{code}}$.

When can we reconstruct the diary from the Hawking radiation? As discussed in section 2, for any pure initial black hole state, there is a phase transition in the Ryu-Takayanagi surface at the Page time. After this point, the diary will be in the entanglement wedge of the Hawking reservoir $\mathcal{H}_{\text{rad}}$ (assuming it was thrown into the black hole at least one scrambling time into the past).

Unfortunately, our lack of knowledge about the state of the black hole prevents us from taking advantage of this fact. Instead, we can only successfully decode the state of the diary once a state-independent reconstruction becomes possible that works for the entire code space of possible initial states.

Such a reconstruction will only be possible once the diary is contained in the entanglement wedge of the reservoir $\mathcal{H}_{\text{rad}}$ even for highly mixed states in the code space of initial microstates (or equivalently code space states that are highly entangled with a reference system $\mathcal{H}_R$). These states will have a large bulk entropy in the interior, which will increase the generalised entropy of the non-empty quantum extremal surface for $\mathcal{H}_{\text{rad}}$, as shown in figure 12. Note that, because we are no longer considering bipartite pure states, the Ryu-Takayanagi surfaces for $\mathcal{H}_{\text{rad}}$ and $\mathcal{H}_{\text{CFT}}$ will no longer necessarily be the same. The Ryu-Takayanagi surface of $\mathcal{H}_{\text{rad}}$ will therefore remain empty for highly mixed states until

$$S_{\text{rad}} - S_{BH} > S_{\text{code}}. \qquad (3.6)$$

Immediately after the Page time, we need to know the exact initial state of the black hole in order to reconstruct the diary. However, as the black hole continues to evaporate, the amount of state-dependence required in the reconstruction decreases; a single reconstruction can work for an increasingly large class of microstates. Happily, (3.6) agrees exactly
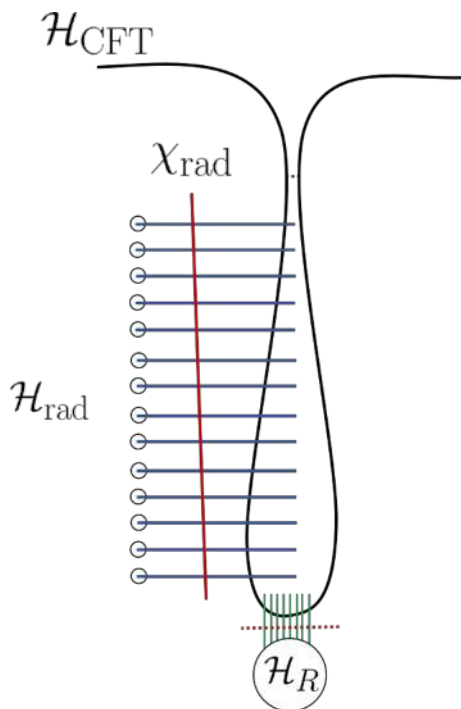
Figure 12. After the Page time, the entanglement wedge of the reservoir contains most of the interior for any pure initial black hole microstate. However, if we only know that the initial state was in some large class of possible microstates, we cannot take advantage of this fact to do a Hayden-Preskill recovery, unless the interior is also contained in the entanglement wedge for states in the code space of possible microstates that are highly entangled with a reference system. This entanglement entropy increases the generalised entropy of the non-empty quantum extremal surface (dotted lines) and can make the Ryu-Takayanagi surface $\chi_{\rm rad}$ become empty (solid line), preventing us for doing a Hayden-Preskill reconstruction until the black hole has evaporated further.

with the amount of state-dependence required for the Hayden-Preskill decoding criterion in simple random unitary toy models [38].

This state dependence does not just make the entanglement wedge version of Hayden-Preskill compatible with toy models. It also provides the *mechanism* by which information about the initial state of the black hole is able to escape out into the Hawking radiation. The Hawking radiation that escapes the boundary is always entangled with interior modes in the same way, regardless of the initial state of the black hole. However, the way that the interior modes are encoded in $\mathcal{H}_{\rm rad}$ depends on the initial state of the black hole. The Hawking radiation will be purified by a different subsystem of $\mathcal{H}_{\rm rad}$, depending on the initial state of the black hole.[56] As a result, the reservoir $\mathcal{H}_{\rm rad}$, plus the additional Hawking radiation, contains more information about the initial state of the black hole than the reservoir alone. The bulk evaporation is consistent with information escaping,

---

[56]We emphasize that, for any single state, the subsystem that purifies some Hawking quanta is not uniquely defined because the decoding channel used to reconstruct the interior mode on $\mathcal{H}_{\rm rad}$ is not unique. The point here is that there is *no* subsystem that purifies the Hawking quanta for all the possible initial states.

even though the state of the Hawking radiation and the interior mode does not care about the initial microstate of the black hole.

Interestingly, even if the initial microstate is completely unknown, and so the code space entropy S_{code} is equal to the initial Bekenstein-Hawking entropy of the black hole, (3.6) will be satisfied long before the black hole has completely evaporated, because of the thermodynamic entropy increase from the evaporation. The information in the diary, as well as all the information about the initial state of the black hole, will be revealed, in a completely state-independent way, even while the black hole is still an $O(1)$ fraction of its original size. This is closely related to the fact that, even for a completely thermal initial black hole state, the von Neumann entropy of the reservoir will peak and begin decreasing, even while the black hole is still an $O(1)$ fraction of its initial size, as discussed by Page in [7]. For black holes in our universe, both events occur when the black hole has approximately 90% evaporated [7].

The same effect happens with reconstructions of the interior on the boundary CFT, before the Page time. In this case, a large amount of bulk entropy in the interior will increase the generalised entropy of the empty surface for $\mathcal{H}_{\text{CFT}}$, and so can make the Ryu-Takayanagi surface for $\mathcal{H}_{\text{CFT}}$ become non-empty. As shown in figure 13, the Ryu-Takayanagi surface of $\mathcal{H}_{\text{CFT}}$ will only be empty for highly mixed/entangled states if

$$S_{\text{code}} < S_{BH} - S_{\text{rad}}. \tag{3.7}$$

The interior can therefore only be reconstructed on $\mathcal{H}_{\text{CFT}}$ in a state-independent way for code spaces with entropy S_{code} satisfying (3.7). Initially, we don't need to know very much about the state of the black hole to reconstruct the entire interior on the boundary CFT. However, as the black hole evaporates, the reconstruction becomes more and more state-dependent. Eventually, just before the Page time, one needs to know the exact initial state of the black hole. Again, (3.7) agrees with toy models.

The part of the interior that lies between the non-empty quantum extremal surface and the boundary is always encoded in the CFT in a completely state independent way, both before and after the Page time. No amount of bulk entanglement with a reference system can stop the entanglement wedge of the CFT from including this region.

The non-empty extremal surface lies at a radius $O(G_N)$ inside the event horizon of the black hole. An outgoing lightcone starting from this extremal surface will therefore hit the singularity after an infalling time equal to the scrambling time plus an $O(\beta)$ correction after the infalling time of the extremal surface. This infalling time is within $O(\beta)$ of the 'current' time, when the last radiation was extracted into $\mathcal{H}_{\text{rad}}$. The worldline of an observer, who jumps into the black hole at an $O(\beta)$ time into the future, will remain entirely within the entanglement wedge of the CFT. The entire experience of the observer in the interior, until the curvature becomes large close to the singularity, will be encoded in the CFT in a completely state-independent way.

This is a somewhat comforting fact. It means that the accessible part of the black hole interior is always encoded in the state of the black hole itself, plus $O(1)$ number of
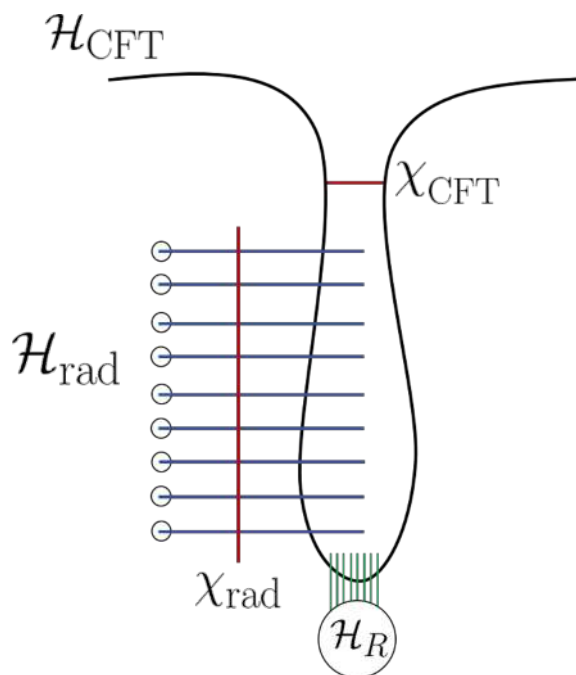
Figure 13. Before the Page time, the interior is in the entanglement wedge of the CFT for any pure state. However, if the initial state is allowed to be highly entangled with a reference system, the Ryu-Takayanagi surface χ_{CFT} for the CFT may jump to the non-empty quantum extremal surface. (In contrast, the RT surface χ_{rad} of the reservoir will remain empty.) This means that a reconstruction of the interior that acts only on the CFT, and not on the reservoir, will have to be at least somewhat state dependent, if the code space is too large.

recent quanta of outgoing Hawking radiation.[57] The initial state of the black hole does not matter, nor do manipulations, even arbitrarily complicated ones, of Hawking radiation that escaped from the black hole a long time in the past.

In particular, so long as $r_s \ll l_{\text{AdS}}$, an observer who jumps into the black hole from the boundary will never leave the entanglement wedge of the CFT at the time that they left the boundary. This seems to still be true, even for large AdS black holes, at least when the Hawking radiation is extracted from inside the zone, as in section 2.2. Potentially, this is important for precomputation versions of the firewall paradox [65], where an observer attempts to extract a mode from $\mathcal{H}_{\text{rad}}$, which is expected to take exponential time [17], before jumping into the black hole.

---

[57]This is somewhat similar to recent work by Yoshida [64], where it was shown, in a qubit toy model of a black hole, that swapping an $O(1)$ number of degrees of freedom and then applying a scrambling unitary was sufficient to make new Hawking radiation be unentangled with the early Hawking radiation, even long after the Page time. Yoshida therefore argued that swapping in the new degrees of freedom had made the interior be encoded in the black hole degrees of freedom, plus the purification of these new degrees of freedom, with no dependence on the initial black hole state. In our case, it is simply the continuous increase with time of the combined thermodynamic entropy of the black hole and Hawking radiation that makes part of the interior be encoded in the CFT. See appendix B for an example of the extremal surface that lies exactly on the event horizon because there is no net increase in thermodynamic entropy.

However, as discussed in appendix B, by making the infalling modes be at a temperature very close to the black hole temperature, we can make the extremal surface be arbitrarily close to the event horizon. By tuning the state of the infalling modes, we can therefore always ensure that the observer is able to escape the entanglement wedge of the CFT and encounter the infalling mode that they extracted into $\mathcal{H}_{\mathrm{rad}}$. The fundamental answer to the precomputation version of the firewall paradox seems to simply be that we can indeed manipulate the interior using $\mathcal{H}_{\mathrm{rad}}$, so long as we are able to do very complicated, non-semiclassical manipulations. After all, it is well known that it is possible to manipulate the interior of a two-sided black hole in the thermofield double state, just by acting on the left CFT.

3.3 Approximation to the rescue

Just like the α-bit codes found in [38] and summarised in section 3.1, the results that we found in section 3.2 only make sense because entanglement wedge reconstruction is approximate. This fact should be somewhat apparent from our discussion in section 3.1 and 3.2. However, in the interests of clarity, we now give a simple, explicit example of a paradox that would otherwise occur.

Suppose that we allow a black hole to evaporate until slightly before the Page time, storing the Hawking radiation in a reservoir $\mathcal{H}_1$. We then allow it to continue to evaporate until slightly after the Page time, storing the Hawking radiation in a different reservoir $\mathcal{H}_2$. Let the entropy of the Hawking radiation in $\mathcal{H}_1$ be $(1-\delta)S_{BH}$ and the entropy of the Hawking radiation in $\mathcal{H}_2$ be $2\delta S_{BH}$. where $\delta > 0$ is small and S_{BH} is the Bekenstein-Hawking entropy of the black hole after all the evaporation has taken place. For reasons that will become clear, we shall refer to the combined state of the evaporating black hole and the Hawking radiation as the 'entangled state'.

The black hole has evaporated beyond the Page time. The Ryu-Takayanagi surface of $\mathcal{H}_{\mathrm{rad}}$ is therefore non-empty and lies just inside the horizon of the black hole. Most of the interior of the black hole is encoded in $\mathcal{H}_1 \otimes \mathcal{H}_2$.

Now suppose that we do a complete measurement of $\mathcal{H}_2$ in some arbitrary basis. Regardless of the outcome of such a measurement, and regardless of the basis that we measure in, the Ryu-Takayanagi surface for the CFT will now be empty and so the interior will be encoded in $\mathcal{H}_{\mathrm{CFT}}$. We shall refer to the states that result from such a measurement as 'pure states', in contrast with the original 'entangled state', because they are pure states in $\mathcal{H}_{\mathrm{CFT}} \otimes \mathcal{H}_1$.

We now have exactly the same apparent paradox that was found in [38] and discussed in section 3.1. For *any* pure state, the interior is encoded in the CFT. However, in the entangled state, it is encoded in $\mathcal{H}_1 \otimes \mathcal{H}_2$ and so, by the no-cloning theorem, it *cannot* be encoded in the CFT. The boundary Hilbert space $\mathcal{H}_{\mathrm{CFT}}$ plays the role of the boundary subregion Hilbert space $\mathcal{H}_B$ that we had access to in section 3.1; the early Hawking radiation reservoir $\mathcal{H}_1$ plays the role of the complementary boundary subregion Hilbert space $\mathcal{H}_{\bar{B}}$ and the later Hawking radiation reservoir $\mathcal{H}_2$ plays the role of the reference system $\mathcal{H}_R$.

As before, the resolution of this paradox is two-fold. Firstly, the reconstruction of the interior on $\mathcal{H}_{\mathrm{CFT}}$ necessarily depends on the state of the interior modes that were previously

entangled with $\mathcal{H}_2$. Hence, there is no single reconstruction that will work for all possible states of those interior modes. This would be necessary to reconstruct the interior of the entangled state in $\mathcal{H}_{\mathrm{CFT}}$.

However, this resolution only works because the reconstruction is approximate. Since the reduced density matrix of the entangled state on $\mathcal{H}_2$ consists of a large number of approximately independently and identically distributed thermal modes, the smooth max entropy of $\mathcal{H}_2$ in the entangled state is approximately equal to its von Neumann entropy $2\delta S_{BH}$, up to subleading corrections of order $O(\sqrt{S})$ [66]. By the definition of the smooth max entropy, this means that we can construct a code space $\mathcal{H}_{\mathrm{code}} \subseteq \mathcal{H}_{\mathrm{CFT}} \otimes \mathcal{H}_1$ satisfying

$$\log |\mathcal{H}_{\mathrm{code}}| = 2\delta S_{BH} + O(\sqrt{S_{BH}}), \tag{3.8}$$

such that the entangled state can be approximated, with very high fidelity, by a state in $\mathcal{H}_{\mathrm{code}} \otimes \mathcal{H}_2$.

Any interior operator $\mathcal{O}$ should have a state-independent global reconstruction $\mathcal{O}^{\mathrm{code}}$ on $\mathcal{H}_{\mathrm{CFT}} \otimes \mathcal{H}_1$ that works for the entire code space $\mathcal{H}_{\mathrm{code}}$. If entanglement wedge reconstruction were exact, then, for any state $|\psi\rangle$, the state-dependent reconstruction $\mathcal{O}^\psi_{\mathrm{CFT}}$ that acts only on the CFT would satisfy

$$\mathcal{O}^\psi_{\mathrm{CFT}} |\psi\rangle = \mathcal{O}^{\mathrm{code}} |\psi\rangle. \tag{3.9}$$

However, as shown in [62, 67], this would imply that there must also exist a state-independent reconstruction $\mathcal{O}^{\mathrm{code}}_{\mathrm{CFT}}$ that works for the entire code space and acts only on the CFT. The cloning paradox can only be resolved if there is an error term in (3.9) with size at least $\exp(-O(\delta\, S_{BH}))$ [62].

This error is tiny; it is non-perturbatively small in G_N for any fixed $\delta > 0$. It is expected that non-perturbative $\exp(-O(S))$ corrections to the bulk physics of black holes in AdS/CFT must exist in order for the decay of correlators to be consistent with boundary unitarity [68]. However, there has been debate about whether such tiny errors can explain the large-scale "$O(1)$" paradoxes that show up in evaporating black holes. The answer is that they can and do. If the code space of allowed microstates has exponentially large dimension, exponentially small errors can be amplified in very entangled states and become $O(1)$ in size.

Once we have measured $\mathcal{H}_2$, regardless of the measurement outcome we obtain, a diary in the interior of the black hole will be encoded in $\mathcal{H}_{\mathrm{CFT}}$ and only has a non-perturbatively small effect on the state of $\mathcal{H}_1$. However, the entanglement with $\mathcal{H}_2$ amplifies these tiny differences, so that orthogonal diary states are almost exactly orthogonal on $\mathcal{H}_1 \otimes \mathcal{H}_2$. No magic is required; just the same mechanism that occurs in random unitary toy models [62].

3.4 Large diaries

So far we have assumed that any diaries thrown into the black hole are small, both in energy and entropy. We have therefore been able to ignore both their backreaction on the geometry and their contribution to the bulk entropy. In this section, we remove those assumptions.

If a heavy diary is thrown into a black hole before the Page time, it can still be reconstructed immediately after the Page time (so long as the entropy of the diary is small). The only change is that the Page time will be delayed by the increase in the horizon area of the black hole caused by the diary. By almost identical arguments to those in section 3.2, if the diary also has a large entropy S_{diary}, we have to wait until

$$S_{\text{rad}} - S_{BH} \geq S_{\text{diary}}, \tag{3.10}$$

so that the diary is contained in the entanglement wedge of $\mathcal{H}_{\text{rad}}$ even for highly mixed diary states.

A more interesting situation occurs when a large diary is thrown into the black hole after the Page time. Let us first consider the case where the entropy of the diary is small, but the energy is large. The diary now causes a large backreaction on the geometry that significantly increases the horizon area. We assume, for simplicity, that the diary is encoded only in s wave modes and so the rotational symmetry of the spacetime is preserved. After we throw the diary into the black hole, the Bekenstein-Hawking entropy of the black hole will again be significantly larger than its entanglement entropy; heuristically, we will have made the black hole young again.

What happens to the quantum extremal surface in such a spacetime? Before the diary is thrown into the black hole, the quantum extremal surface will continuously move forwards in infalling time, at a radius just inside the event horizon, as radiation escapes the black hole into $\mathcal{H}_{\text{rad}}$. However, radiation that would have escaped shortly after the diary was thrown into the black hole will instead fall into the larger black hole created by the backreaction of the diary.

The actual radiation that escapes instead comes from close to the new event horizon, which, being teleological, already began moving out from the apparent horizon r_s along an outgoing lightcone in anticipation of the diary falling in. Even at very late times, the Hawking radiation comes from outside the past lightcone of the boundary, which in turn will always lie outside the event horizon, although it will approach the horizon exponentially at any fixed time as we evolve the boundary forwards in time.

Using (2.31) and (2.34), it is therefore easy to see that the quantum extremal surface stops tracking along the horizon after the diary is thrown into the black hole, and instead asymptotes to a radius

$$r = r_s - \frac{G_N c_{\text{evap}}}{3(d-1)\Omega_{d-1}r_s^{d-2}}, \tag{3.11}$$

at the infalling time v when

$$r_{\text{hor}}(v) = r_s + \frac{G_N c_{\text{evap}}}{3(d-1)\Omega_{d-1}r_s^{d-2}}. \tag{3.12}$$

For simplicity, we are assuming here, as in sections 2.2 and 2.3, that the Hawking radiation is extracted from close to the horizon and so there are no greybody factors.

If the change in horizon area δA_{diary} caused by the diary is small compared to the original horizon area A_{hor}, we find that

$$v = v_{\text{diary}} - \frac{\beta}{2\pi} \log \frac{\delta A_{\text{diary}}}{c_{\text{evap}}G_N} + O(\beta), \tag{3.13}$$

where v_{diary} is the infalling time at which the diary is thrown into the black hole. In deriving (3.13), we have used the fact that the event horizon is an outgoing lightcone, obeying (2.10), and that, at v_{diary}, the radius of the event horizon should be approximately equal to the new Schwarzschild radius of the black hole.

The entanglement wedge of $\mathcal{H}_{\text{rad}}$ will not contain the diary. The large amount of energy thrown into the black hole has stopped any information from escaping. The location of the extremal surface and the entanglement wedges in shown in figure 14.

However, this quantum extremal surface will not remain the Ryu-Takayanagi surface forever. As new radiation escapes from the black hole into $\mathcal{H}_{\text{rad}}$, the bulk entropy, calculated using this quantum extremal surface, will increase. Heuristically, the new Hawking radiation is entangled with interior modes that lie close to the new, larger black hole horizon and hence lie in the entanglement wedge of $\mathcal{H}_{\text{CFT}}$. More formally, as in (2.42), the cut-off ε_2 in (2.19) shrinks exponentially as

$$\varepsilon_2 \propto \exp(-2\pi v_{\text{rad}}/\beta) \tag{3.14}$$

where v_{rad} is the 'current' infalling time (i.e. the point in time where we last extracted Hawking radiation into $\mathcal{H}_{\text{rad}}$). In contrast, the radial distance

$$r_{l.c.} - r \approx r_{hor} - r, \tag{3.15}$$

between the past lightcone and the RT surface is approximately constant. As a result, we find using (2.19) that

$$\frac{\partial S_{\text{bulk}}}{\partial v_{\text{rad}}} = \frac{c_{\text{evap}}\pi}{6\beta}. \tag{3.16}$$

This is the increase in entropy that one finds with thermal outgoing modes that are purified by degrees of freedom in $\mathcal{H}_{\text{CFT}}$ [57, 58]. By adding energy, we have stopped information escaping the black hole and made the Hawking radiation be purely entangled with the CFT. The black hole did indeed become young again.

The surface χ_1 at (3.11), (3.13) remains a quantum extremal surface even at boundary times long after the diary was thrown into the black hole. However, it is not the only non-empty quantum extremal surface at such late times. There will also be a second non-empty quantum extremal surface χ_2 that lies, as usual, approximately the scrambling time before the current boundary time. Because of the increase in horizon area created by the diary, the generalised entropy of the extremal surface χ_2 will initially be significantly larger than the generalised entropy of the surface χ_1.

However the generalised entropy of χ_1 steadily increases over time because of the increase in bulk entropy discussed above. Meanwhile, the generalised entropy of χ_2 decreases over time as the black hole evaporates. Eventually, after the new Page time of the black hole, the generalised entropy of χ_2 will become smaller than the generalised entropy of χ_1. The later extremal surface χ_2 will be the Ryu-Takayanagi surface, and the diary will finally be in the entanglement wedge of $\mathcal{H}_{\text{rad}}$ and can be decoded.

If the diary had a large amount of entropy, as well as a large amount of energy, we would have to wait even longer in order to recover the diary. Even after the new Page time,
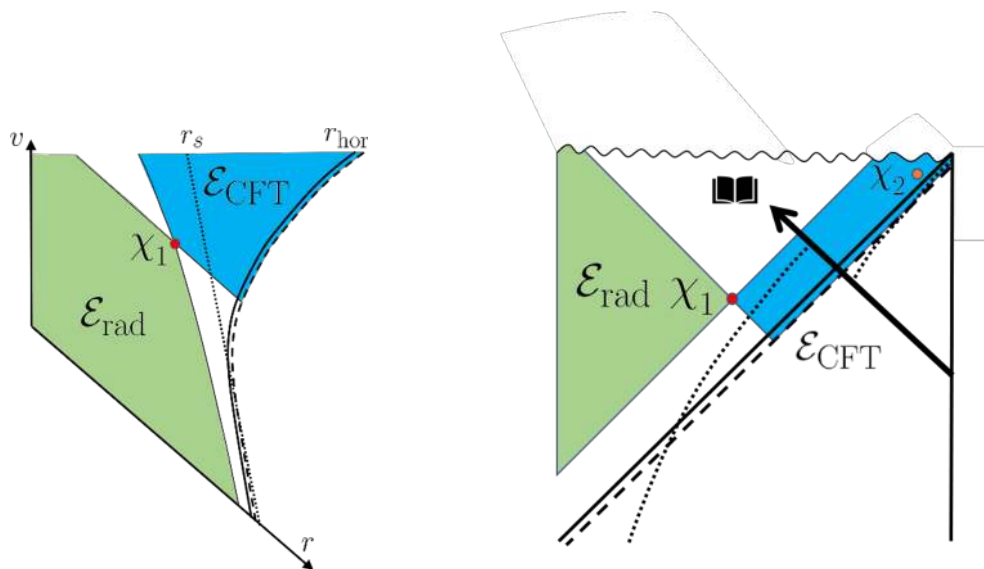
Figure 14. When a large diary is thrown into a black hole, the radius r_{hor} of the event horizon (solid line) begins increasing in anticipation of the diary falling in, while the radius r_s of the apparent horizon (dotted line) continues to slowly decrease until the diary is actually thrown into the black hole. As the black hole continues to evaporate, the past lightcone (dashed line) of the current boundary remains outside the event horizon. The Ryu-Takayanagi surface χ_1 asymptotes to a point approximately the scrambling time before the diary was thrown in. There is a second quantum extremal surface χ_2 at an infalling time after the diary is thrown in. Initially, this extremal surface is not the RT surface because its area is larger than the area of χ_1. However, eventually, at the new Page time, there will be a phase transition, with χ_2 becoming the new RT surface, and the diary can finally be reconstructed from the Hawking radiation, so long as its entropy is small. (Left: Eddington-Finkelstein coordinates, right: a Penrose diagram.)

when the Ryu-Takayanagi surface has a phase transition for pure diary states, the Ryu-Takayanagi surface of $\mathcal{H}_{\text{rad}}$ will still not contain the diary for highly mixed diary states. To actually recover the diary from the Hawking radiation, we need to wait until the diary is contained in the entanglement wedge of $\mathcal{H}_{\text{rad}}$, even for such highly mixed states, as shown in figure 15. This requires

$$S_{\text{rad}}^{\text{new}} - \frac{\delta A_{\text{diary}} + \delta A_{\text{evap}}}{4 G_N} > \log |\mathcal{H}_{\text{diary}}| \tag{3.17}$$

where $S_{\text{rad}}^{\text{new}}$ is the bulk entropy of the new radiation emitted after the diary was thrown into the black hole, $\delta A_{\text{diary}} > 0$ is the change in horizon area from throwing the diary into the black hole, $\delta A_{\text{evap}} < 0$ is the change in horizon area from the black hole evaporation after the diary is thrown into the black hole and $\mathcal{H}_{\text{diary}}$ is the Hilbert space of the diary.

Note that the generalised second law implies that

$$\delta A_{\text{diary}}/4 G_N \geq \log |\mathcal{H}_{\text{diary}}| \tag{3.18}$$

and

$$S_{\text{rad}}^{\text{new}} \geq -\delta A_{\text{evap}}. \tag{3.19}$$

Hence, whenever (3.17) is satisfied, we will also have

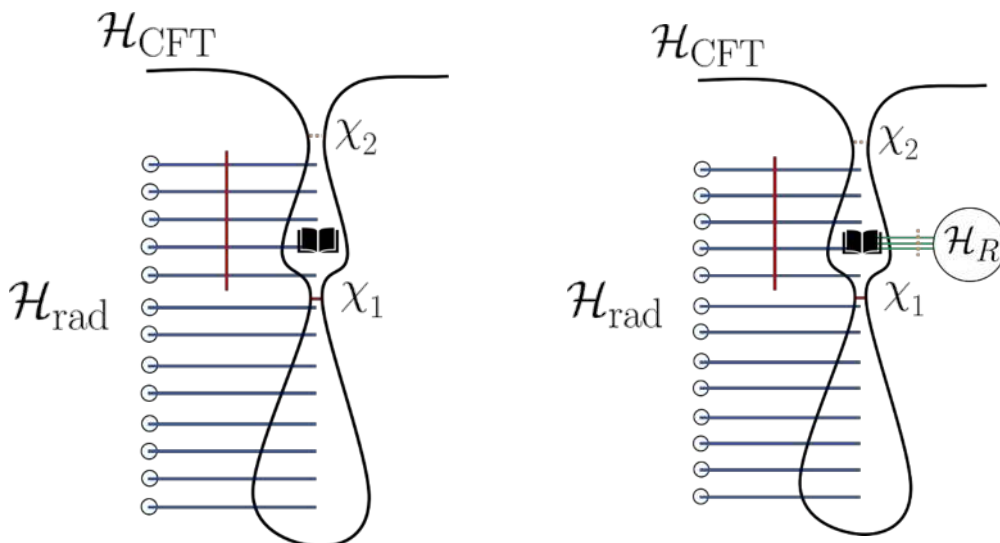$$S_{\text{rad}}^{\text{new}} > \log |\mathcal{H}_{\text{diary}}|. \tag{3.20}$$

Figure 15. If a heavy diary, left, is thrown into the black hole, it can only be reconstructed from $\mathcal{H}_{\mathrm{rad}}$ once the generalised entropy (solid lines) of the surface χ_1 for $\mathcal{H}_{\mathrm{rad}}$ is greater than the generalised entropy (dotted lines) of the surface χ_2. The generalised entropy of χ_1 has contributions from both the area term and the entropy of Hawking radiation emitted after the diary was thrown into the black hole. In contrast, so long as the diary has small entropy, the surface χ_2 only has a contribution from its area, plus an $O(1)$ bulk entropy correction. However, if the diary also has a large entropy (right), the diary also needs to be in the entanglement wedge, even when it is in a highly mixed state, or is entangled with a reference system. This increases the generalised entropy of χ_2.

It follows that there is always sufficient entropy in the new Hawking radiation to encode the diary. Entanglement wedge reconstruction is consistent with quantum capacity bounds, so long as we consider mixed (or entangled) states in the code space.

It can be verified (3.17) is consistent with random unitary toy models [38], although, as discussed in section 2.3, most of the focus in random unitary models has been on evaporation that is either perfectly or approximately thermodynamically reversible, where (3.18) and (3.19) are equalities.

As in section 3.2, the period when the interior reconstruction depends on the state of the diary does not merely make the amount of information encoded in the Hawking radiation be compatible with quantum capacity bounds; it also provides the mechanism by which information about the state of the diary escapes the black hole. The new Hawking radiation is entangled with the same interior modes, regardless of the state of the diary. However, because the encoding of those interior modes in $\mathcal{H}_{\mathrm{rad}}$ depends on the state of the diary, an observer with access to the reservoir $\mathcal{H}_{\mathrm{rad}}$ can learn information about the diary from the new Hawking radiation.

So far, both in this section and in section 2.3, we have not worried too much about the errors that exist in entanglement wedge reconstruction, even though we showed in section 3.3 that their existence was crucial to the consistency of our results. Reconstruction errors are also present in the Hayden-Preskill protocol in random unitary toy models of black holes. Since all our other results have been consistent with random unitary toy mod-

els, we might hope that the error in the Hayden-Preskill entanglement wedge reconstruction will also be consistent with random unitary toy models.

Unfortunately, the actual size of the errors in entanglement wedge reconstruction remains unknown. However, the *lower bound* on their size that was derived in [38] and discussed briefly in section 3.1 suggests that, for reconstruction with error ε to be possible, the difference ΔS between the generalised entropy of an extremal surface for which the bulk operator would not be in the entanglement wedge and the generalised entropy of the Ryu-Takayanagi surface, where the bulk operator is in the entanglement wedge, must satisfy

$$\Delta S \geq O\left(\log \frac{1}{\varepsilon}\right). \tag{3.21}$$

Note that (3.21) needs to be satisfied for all states, both pure and mixed, in the code space. The exact coefficient in (3.21) depends on how the error is measured, and we will not worry about it here.

The generalised entropy of the extremal surface χ_1, where the diary is not in the entanglement wedge of $\mathcal{H}_{\rm rad}$, increases by an $O(1)$ amount in $O(\beta)$ time. Similarly, the generalised entropy of the extremal surface χ_2, where the diary is in the entanglement wedge of $\mathcal{H}_{\rm rad}$, decreases by an $O(1)$ amount in $O(\beta)$ time.

If we make the strong assumption that the lower bound on the error ε derived in [38] is approximately saturated, we find that to reconstruct the diary with error ε, we need to wait for an additional time

$$O(\beta \log \left(\frac{1}{\varepsilon}\right),$$

even after the condition (3.17) is satisfied. Up to the (unstated) linear coefficient, this agrees, yet again, with random unitary toy models [8, 62].

3.5 Minimal state dependence

So far, we have avoided talking about state dependence for interior operators in black holes that have not evaporated at all, where there is no auxiliary reservoir $\mathcal{H}_{\rm rad}$. Yet this is the situation in which state dependence is most commonly discussed [32, 33, 37].

There is a very good reason for our reticence. Every proof of entanglement wedge reconstruction assumes that there is a global isometry from the bulk code space to the larger boundary Hilbert space. It is this isometry, combined with a partial trace over some of the boundary degrees of freedom, that creates a noisy quantum channel and, potentially, a quantum error correcting code. Yet, so long as such an isometry exists, there must always exist state-independent global boundary reconstructions.

If the CFT is the only Hilbert space and pure bulk states correspond to pure boundary states, interior operators cannot be encoded in the CFT in a state-dependent way, within the framework of quantum error correction.

Nonetheless, suppose we take as an axiom the idea that boundary reconstructions are state independent if, and only if, the bulk operator is contained in entanglement wedge even for states in a 'code space' of the bulk effective field theory that are entangled with a

reference system.[58] If a code space of interior states $\mathcal{H}_{\text{code}}$ satisfies

$$\log |\mathcal{H}_{\text{code}}| < S_{BH}, \tag{3.22}$$

the entanglement wedge of the boundary should always include the interior, even if the state is entangled with a reference system.[59] A single reconstruction of interior operators should therefore exist that works for the entire code space.

In contrast, if

$$\log |\mathcal{H}_{\text{code}}| \geq S_{BH}, \tag{3.23}$$

then we can make states in $\mathcal{H}_{\text{code}} \otimes \mathcal{H}_R$ where the RT surface is non-empty, and part of the interior is no longer contained in the entanglement wedge of the boundary.

To distinguish this idea from most of the literature on interior state dependence, which has focussed on constructing interior operators for a single microstate, or at most a small code space of microstates with $O(1)$ dimension, we shall refer to it as 'minimal state dependence'. We emphasize, however, that this paper is far from the first to suggest it, see, for example, the discussion near the end of [32].

As discussed, given that we are taking results derived using quantum error correction and applying them outside of that framework, we should be somewhat cautious about this idea. In particular, rather than making too many claims about code spaces for which (3.23) is true, it seems better to focus on the fact that, so long as a code space satisfies (3.22), we should be relatively confident that everything is well behaved and that all bulk operators have global, state-independent boundary reconstructions that work for the entire code space.

An important question is how small

$$\Delta S = S_{BH} - \log |\mathcal{H}_{\text{code}}| \tag{3.24}$$

can be without causing any problems. However, since we don't have the tools to answer such a question with any confidence, we shall be maximally cautious and assume that ΔS needs to be non-zero at leading order. In other words, we shall require

$$\log |\mathcal{H}_{\text{code}}| \leq (1-\delta)S_{BH}, \tag{3.25}$$

for some $\delta > 0$, which should be fixed in the semiclassical limit.

We can now show how minimal state dependence neatly avoids the so-called AMPSS argument [65] that generic black hole microstates must have firewalls.

The AMPSS argument goes as follows. Consider the subspace of CFT states within some narrow, but $O(1)$ width, energy band $M \leq E \leq M + \delta M$. If M is sufficiently large and δM is sufficiently small then all such states have a bulk description as a black hole.

---

[58]We emphasize that this may no longer be a code space of a quantum error correcting code in the traditional sense.

[59]The dimension of the code space here includes both allowed interior degrees of freedom, as well as any degrees of freedom describing an end-of-the-world brane that are also included in the code space, if our geometry ends in such a brane. See, for example, [37, 69] for discussion of interior geometries ending on end-of-the-world branes.

We then assume that there exists some state-independent operator $b_\omega^\dagger$ that acts as a raising operator for an interior Hawking mode, as well as an inverse operator $(1 + b_\omega^\dagger b_\omega)^{-1} b_\omega$.

Acting with the operator $b_\omega^\dagger$ decreases the Schwarzschild energy by ω and so maps our subspace of CFT states into the energy band $M - \omega \le E \le M + \delta M - \omega$. But the number of states in this energy band is smaller than the number in our original energy band by a factor of approximately $e^{-\beta\omega}$. So the map cannot be invertible, contradicting our original assumption. The authors of AMPSS conclude that $b_\omega^\dagger$, and hence the interior, cannot exist.

It is, of course, well known that the argument above breaks down if the operator $b_\omega^\dagger$ is state dependent. Our point here is simply to emphasize that this is still true even if the operator $b_\omega^\dagger$ is only *minimally* state dependent.

Up to logarithmic corrections, the entropy of the maximally mixed state in the energy band $M \le E \le M + \delta M$ is equal to the Bekenstein-Hawking entropy. Assuming minimal state-dependence, we won't be able to find a single state-independent operator $b_\omega^\dagger$. Our code space is too large. If instead we restrict to a random code subspace, within the energy band and with entropy at most $(1 - \delta)S_{BH}$, there will be plenty of space for the image of the code subspace in the energy band $M - \omega \le E \le M + \delta M - \omega$.

Of course, instead of looking at a random subspace of states within the energy window, we can alternatively make δM so small that the entire code space of states in the energy band has a single state-independent reconstruction. To do so, it would be necessary to have

$$\delta M = O(e^{-\delta S_{BH}}). \tag{3.26}$$

If we insisted that the energy band $M \le E \le M + \delta M$ be mapped invertibly into the energy band $M - \omega \le E \le M + \delta M - \omega$, we would still have a paradox, because the latter band is still strictly smaller than the former.

This would be far too strong a requirement however. More reasonably, we should only expect the energy of a state to decrease by ω plus non-perturbatively small corrections. However, since δM is itself non-perturbatively small, this non-perturbatively small uncertainty in ω can significantly increase the size of the energy window. We can therefore avoid the paradox.

4 Discussion

4.1 Summary of results

In this paper, we have argued that the key expected features of unitary black hole evaporation in AdS/CFT can be derived from the bulk semiclassical description of an evaporating black hole, so long as we assume entanglement wedge reconstruction. We now review those arguments.

- We studied entanglement wedge reconstruction in an evaporating black hole, formed from collapse, where the Hawking radiation was extracted out of the AdS space containing the black hole, with boundary Hilbert space $\mathcal{H}_{\text{CFT}}$, and into an auxiliary Markovian reservoir $\mathcal{H}_{\text{rad}}$. Importantly, we assumed that the bulk description of the

evaporation was semiclassical, and so the bulk entanglement between the Hawking radiation and the interior of the black hole continued to grow, even after the Page time.

- Using the maximin prescription, we saw that the quantum Ryu-Takayanagi surface, associated to the entire boundary Hilbert space $\mathcal{H}_{\text{CFT}}$, must become non-empty at the Page time. Since the overall state $|\psi\rangle \in \mathcal{H}_{\text{CFT}} \otimes \mathcal{H}_{\text{rad}}$ is pure, this non-empty RT surface will also be the RT surface for $\mathcal{H}_{\text{rad}}$.

- If the Hawking radiation is extracted into $\mathcal{H}_{\text{rad}}$ from deep inside the zone, near the horizon, the greybody factors will all be either zero or one, depending on whether a given angular momentum mode is extracted or not. The location of the non-empty quantum Ryu-Takayanagi surface can then be calculated explicitly. It lies at a radius

$$r = r_s - \frac{c_{\text{evap}} G_N}{3(d-1)r_s^{d-2}\Omega_{d-1}} = r_{\text{hor}} - \frac{c_{\text{evap}} G_N}{6(d-1)r_s^{d-2}\Omega_{d-1}}, \tag{4.1}$$

where r_s is the radius of the classical apparent horizon of the black hole, r_{hor} is the radius of the event horizon of the black hole and c_{evap} is the number of modes that are extracted into the reservoir $\mathcal{H}_{\text{rad}}$. The infalling time of the quantum extremal surface is given by

$$v = -\frac{\beta}{2\pi} \log \frac{S_{BH}}{c_{\text{evap}}} + O(\beta), \tag{4.2}$$

where $v = 0$ is the current boundary time. A large part of the black hole interior lies in the entanglement wedge of the Hawking radiation reservoir $\mathcal{H}_{\text{rad}}$, rather than the boundary Hilbert space $\mathcal{H}_{\text{CFT}}$.

- A small diary thrown into the black holes early in the evaporation can therefore be reconstructed from the Hawking radiation immediately after the Page time. A small diary thrown into the black hole after the Page time can be reconstructed after waiting for the scrambling time. These two results constitute the Hayden-Preskill decoding criterion [8].

- If the number of angular momentum modes c_{evap} extracted into the Hawking radiation is large, there is a small, logarithmic decrease in the delay before the diary can be decoded from the radiation. This decrease is consistent with a heuristic picture of fast scrambling where a perturbation spreads exponentially through the degrees of freedom.

- More generally, we showed that, in *any* evaporating black hole after the Page time, the RT surface lies at an infalling time

$$v = -\frac{\beta}{2\pi} \log \frac{1}{G_N} + O(\beta), \tag{4.3}$$

where the subleading corrections depend on the details of the evaporation and, in general, cannot be analytically calculated, because of greybody factors. We can therefore derive the Hayden-Preskill decoding criterion, up to unknown, but subleading, corrections, even when non-trivial greybody factors are present.

- Given the location of the Ryu-Takayanagi surface, it is an immediate consequence of the Ryu-Takayanagi formula that the entanglement between the CFT and the reservoir is given by the Page curve. Moreover, entanglement wedge reconstruction explains *how* the entanglement entropy ends up decreasing (and how the AMPS firewall paradox is avoided). It decreases because the outgoing radiation is entangled with interior modes that are in the entanglement wedge of, and so are encoded in, $\mathcal{H}_{\text{rad}}$.

- If we consider the change in bulk entanglement entropy from transferring a small amount of Hawking radiation from $\mathcal{H}_{\text{CFT}}$ to $\mathcal{H}_{\text{rad}}$, we find exact quantitative agreement with the change in entropy given by the Ryu-Takayanagi formula, i.e. the Page curve.

- This quantitative agreement does not only exist in the simple cases where we can calculate the Ryu-Takayanagi surface explicitly. It is a general consequence of the fact that the quantum Ryu-Takayanagi surface is an extremum of the generalised entropy. As with the Hayden-Preskill decoding criterion, we are therefore able to derive the Page curve, and avoid the firewall paradox, even when there are greybody factors.

- As argued in [38], based on results about approximate operator algebra quantum error correcting codes derived in [39–41], *state-independent* entanglement wedge reconstruction is only possible for a given code space if the bulk operator is contained in the entanglement wedge of the boundary region for all states, *both pure and mixed* that have support only within the code space.

- *State-dependent* entanglement wedge reconstruction is possible so long as the bulk operator is contained in the entanglement wedge of the boundary region for all *pure states* in the code space.

- Using these results, we were able to derive the correct state dependence for Hayden-Preskill reconstructions. Immediately after the Page time, the diary can only be reconstructed if the exact initial black hole microstate is known. As the black hole continues to evaporate, less state dependence is required, in exact agreement with toy models. Specifically, a single reconstruction will work for a large code space of possible initial black hole states, with entropy S_{code}, so long as

$$S_{\text{code}} < S_{\text{rad}} - S_{BH}, \tag{4.4}$$

where S_{rad} is the bulk entanglement entropy between the Hawking radiation and the interior and S_{BH} is the Bekenstein-Hawking entropy of the black hole.

- Similarly, before the Page time, reconstructions of the interior on the boundary CFT become increasingly state dependent as the black hole evaporates. The entropy S_{code} of the code space of allowed initial states must satisfy

$$S_{\text{code}} < S_{BH} - S_{\text{rad}}. \tag{4.5}$$

Eventually, at the Page time, the reconstruction is only possible if the exact initial black hole state is known.

- The state dependence in the encoding of interior partners of Hawking modes in the early radiation (after the Page time) explains how the final (microscopic) state of the combined early and late radiation can depend on the initial state of the black hole (i.e. how we can avoid information loss) despite the state of the late Hawking mode and its interior partner having no dependence on the initial state.

- These results are only consistent because entanglement wedge reconstruction is only approximate. Tiny, non-perturbatively small errors build up and have $O(1)$ effects.

- When a heavy diary is thrown into the black hole, the Ryu-Takayanagi surface stops tracking along the horizon and instead asymptotes to a location approximately one scrambling time before the diary was thrown in. The Hawking radiation will contain no further information until the new Page time is reached, at which point the Ryu-Takayanagi surface will jump forwards in time and the diary can be decoded from the Hawking radiation.

- If the *entropy* of the diary is large, as well as its energy, we have to wait even longer before it can be decoded. Specifically, we have to wait until the generalised entropy of the earlier quantum extremal surface, where the diary is not in the entanglement wedge of $\mathcal{H}_{\mathrm{rad}}$, is greater than the generalised entropy of the later quantum extremal surface plus the entropy of the diary code space.

- All our results about Hayden-Preskill reconstructions of large diaries are consistent with random unitary toy models [38].

- If we assume that the lower bound, derived in [38], on errors in entanglement wedge reconstruction is saturated up to a linear coefficient, we find that the errors in Hayden-Preskill reconstructions are consistent with toy models.

- Finally, our arguments suggest that, even when a black hole has not evaporated at all, its interior can only be reconstructed with 'minimal' state dependence. Such state dependence is beyond the framework of quantum error correction, but it provides a natural resolution to the AMPSS typical state firewall paradox.

In appendices,

- We give a simple pedagogical example of the importance of the coordinate dependence of cut-offs in entanglement entropy calculations.

- We calculate the location of the Ryu-Takayanagi surface explicitly when the infalling modes are replaced by thermal modes or by pure modes of constant energy density. In each case we find that information stops escaping in the Hawking radiation (from the perspective of an observer with access either only to the reservoir, or to the reservoir and a purification of the thermal infalling modes) exactly at the moment that no information escaping becomes consistent with boundary unitarity.

- We show that the Kourkoulou-Maldacena state-dependent interior reconstruction in the SYK model can be trivially extended to give minimally state-dependent reconstructions.

4.2 Entanglement wedge reconstruction in toy models

Throughout this paper, we have found that bulk calculations, using entanglement wedge reconstruction and the Ryu-Takayanagi formula, agree perfectly with random unitary and fast scrambling toy models of the boundary dynamics.

Not only do our results agree with the Page curve and the Hayden-Preskill decoding criterion, but we also found exactly the right state dependence, for both reconstructions acting on the CFT before the Page time and reconstructions acting on the reservoir after the Page time. Our results for large diaries were consistent with toy models as a function of the both the energy and the entropy of the diary, as were the reconstruction errors so long as we assumed that the lower bound from [38] was approximately saturated.

This seems either to be a somewhat remarkable coincidence, or to involve some deep magic of quantum gravity. In fact, it is neither of these things. We simply have the direction of causation in reverse. Rather than random unitary models determining the consequences of entanglement wedge reconstruction, entanglement wedge reconstruction determines the behaviour of random unitary toy models.

A random unitary toy model of black hole evaporation is an exceptionally trivial example of a random tensor network [70]. An iterated random isometry model like that in figure 9 is another, more complicated, example.

However, random tensor networks are well known to obey the Ryu-Takayanagi formula and hence have entanglement wedge reconstruction [70]. It is therefore inevitable that random unitary models agree with results derived *from* Ryu-Takayanagi and entanglement wedge reconstruction. Indeed, one of the most popular methods to prove results about error correction in random unitaries, the so-called decoupling approach [71, 72], essentially involves deriving entanglement wedge reconstruction from the Ryu-Takayanagi formula.

Of course, random tensor networks do not have all the properties of holographic spacetimes. In particular, they are not covariant. A tensor network corresponds to a single timeslice of a bulk spacetime; RT surfaces, in a tensor network, are minimal, not extremal, surfaces.

For many of the calculations in this paper, for example the scrambling time delay in the Hayden-Preskill criterion, the extremality of the surface, or equivalently the maximisation over Cauchy slices in the maximin prescription, was crucial in deriving the correct results. In particular, the covariant Ryu-Takayanagi surface somehow knows about the fast scrambling dynamics of the boundary theory. If there is any magic going on, it seems to be here.

4.3 The post evaporation state and the bulk-to-boundary map

While we have studied evaporating black holes both before and after the Page time in this paper, we have not discussed the final state of the system, after the evaporation is complete. We make some comments about this state now.
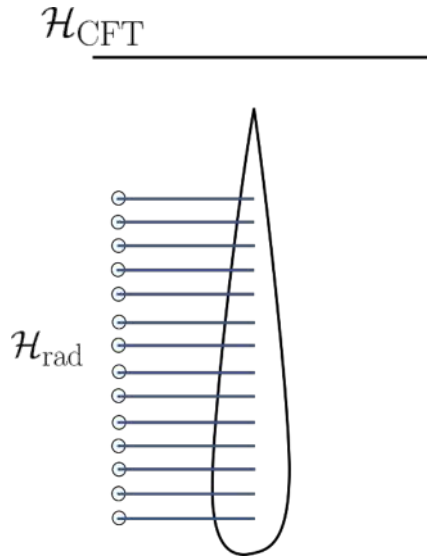
Figure 16. After the black hole has completely evaporated, the bulk encoded in the CFT will be in the vacuum state. However, we can choose a (disconnected) bulk Cauchy slice where the radiation in the reservoir $\mathcal{H}_{\mathrm{rad}}$ is still be entangled with the pinched-off interior wormhole, which lies in its entanglement wedge. The Ryu-Takayanagi surface is empty and has zero generalised entropy; the two systems $\mathcal{H}_{\mathrm{CFT}}$ and $\mathcal{H}_{\mathrm{rad}}$ are therefore unentangled.

When the black hole has nearly completely evaporated, the horizon curvature will become large and so stringy and Planckian effects will become important. We can no longer trust semi-classical calculations.

However it is reasonable to expect that, after an indeterminate, but short, time, there will no longer be any sort of smooth connected geometry between the wormhole and the AdS space that previously contained the black hole. The mouth of the wormhole will have closed; the black hole will have completely evaporated. Let us assume we have extracted all the remaining energy out of the AdS space and into the reservoir $\mathcal{H}_{\mathrm{rad}}$, so that the original bulk AdS space lies in the vacuum state. This is shown schematically in figure 16.

Because the closed wormhole geometry has no boundary, on its own, the Ryu-Takayanagi *surface* is no longer sufficient to define an entanglement wedge, and hence a generalised entropy. For example, if the RT surface is empty, we need to further specify whether the wormhole is in the entanglement wedge of the reservoir or of the CFT, before the generalised entropy is well-defined. In this case, it is obvious that the RT surface should be empty, with the wormhole in the entanglement wedge of the reservoir, since this gives zero generalised entropy.[60]

There is no entanglement between $\mathcal{H}_{\mathrm{CFT}}$ and $\mathcal{H}_{\mathrm{rad}}$; both states are pure. Entanglement wedge reconstruction tells us that the state of the wormhole is encoded in the Hawking radiation. Any information thrown into the black hole during the evaporation will be contained in the entanglement wedge of $\mathcal{H}_{\mathrm{rad}}$, no matter how large the entropy of the

---

[60]Recall that the empty surface is contained in every Cauchy slice, and so will be the RT surface if there is any Cauchy slice for which it has minimal generalised entropy.

initial state. The information thrown into the black hole is therefore encoded in $\mathcal{H}_{\mathrm{rad}}$ in a completely state-independent way.[61] All the information has been preserved.

Even though we know from the Ryu-Takayanagi formula that the state of the Hawking radiation in $\mathcal{H}_{\mathrm{rad}}$ is pure, it appears from a bulk perspective that it is still entangled with the closed-off wormhole. How should we understand this seeming contradiction?

An entangled state of the Hawking radiation and wormhole can be written as

$$|\psi\rangle = \sum_i \sqrt{p_i}\, |\phi_i\rangle\, |\chi_i\rangle\,, \tag{4.6}$$

where the states $|\phi_i\rangle$ describe the radiation and the states $|\chi_i\rangle$ describes the wormhole. The Hilbert space of a holographic theory is associated with its boundary. However, the wormhole has no boundary. We therefore conclude (perhaps somewhat controversially) that its Hilbert space is trivial; it is isomorphic to the complex numbers $\mathbb{C}$. The states $|\chi_i\rangle$ are therefore simply complex coefficients c_i. The state

$$|\psi\rangle = \sum_i \sqrt{p_i}\, c_i\, |\phi_i\rangle \tag{4.7}$$

is therefore simply some complicated, pure state in $\mathcal{H}_{\mathrm{rad}}$.

Of course, there exists a perfectly valid bulk description of the same bulk state $|\psi\rangle$ from (4.7) where the Hawking radiation is simply in a bulk state that has a complicated entanglement structure, but no wormhole.[62] Some version of black hole complementarity, or the $ER = EPR$ duality [18], makes it equally valid to suppose that the wormhole still exists, or that it has vanished leaving some complicated pure state of the Hawking radiation.

If we apply some complicated unitary operator to $\mathcal{H}_{\mathrm{rad}}$, we can transform the complicated state $|\psi\rangle$ into a simple state, say $|\psi_0\rangle$. From the perspective where the wormhole continues to exist, we will still be in an 'entangled state'

$$\sum_i \sqrt{p_i'}\, |\psi_i\rangle\, |\chi_i'\rangle\,, \tag{4.8}$$

where there are many non-zero p_i'. However, the states $|\chi_i'\rangle$ are now very complicated superpositions of the original 'simple' wormhole states $|\chi_i\rangle$. Each term in each superposition is simply a complex number. For $|\chi_0'\rangle$ the coefficients in the superposition must add constructively, so that $|\chi_0'\rangle$ is some large complex number c_0'. For $|\chi_i'\rangle$ with $i \neq 0$, they must interfere destructively so that $|\chi_i'\rangle = 0$. The 'entangled state' really is just $|\psi_0\rangle$.

Just like in ordinary AdS/CFT, we have a linear 'bulk to boundary' map from the state of bulk fields $\mathcal{H}_{\mathrm{bulk}}$ on a fixed background spacetime (in this case the closed wormhole) to a 'boundary' state of the spacetime itself. However, since the 'boundary' Hilbert space is trivial, up to normalisation, this map simply projects the bulk fields on the wormhole into a particular, very complicated state.

---

[61]In contrast, the interior of the black hole, from before it began to evaporate, appears to be encoded in the reservoir $\mathcal{H}_{\mathrm{rad}}$ in exactly the same minimally state dependent way that it was originally encoded in the CFT.

[62]If it is not clear that the Markovian reservoir $\mathcal{H}_{\mathrm{rad}}$ has a bulk description at all, recall that we can imagine throwing each small chunk of Hawking radiation into its own copy of anti-de Sitter space.

The coefficients c_i which describe the 'boundary' state associated to a particular worm-hole state are simply the coefficient of that state in the projected wormhole state. This story is, in effect, simply the Horowitz-Maldacena final state postselection proposal [42]. From a bulk perspective, Horowitz and Maldacena suggested that the projection happens at the singularity. Our arguments suggest that, in the microscopic boundary description of the theory, it happens the moment that the wormhole closes off, even if we choose a bulk Cauchy slice that includes the wormhole.

In fact, the process that ends in this final state projection begins immediately after the Page time, long before the black hole fully evaporates. At this point, there are, for the first time, more 'orthogonal' states of the interior fields, that are needed to describe the entanglement with the Hawking radiation, than there are microstates of the black hole according to the Bekenstein-Hawking entropy.

As a result, even though each 'orthogonal' pair of bulk states will be almost orthogonal on the boundary, there must exist very complicated superpositions of states of the interior fields that are annihilated by the map to the boundary. Indeed, this is what allows the entanglement entropy between the reservoir and the CFT to be much less than the bulk entanglement entropy between the radiation and the interior, in accordance with the Ryu-Takayanagi formula.

As with the post-evaporation state, the combined state of the black hole and radiation, after the Page time can be written as

$$|\psi\rangle = \sum_i \sqrt{p_i} \, |\phi_i\rangle \, |\chi_i\rangle \, , \tag{4.9}$$

where the probabilities p_i are determined by the semiclassical bulk evaporation, the states $|\phi_i\rangle$ describe the Hawking radiation in the reservoir and the states $|\chi_i\rangle$ now describe black hole microstates, which are encoded as CFT states. If there was an isometry, or even an approximate isometry, mapping each apparently distinct microstate $|\chi_i\rangle$ to a different CFT state, the entanglement entropy between the reservoir and the CFT would be equal to S_{rad}. However, minimal state dependence says that the bulk 'code space' of microstates $|\chi_i\rangle$ is too large for such an isometry to exist. Since the map from bulk states to boundary states is not an isometry, there is no inconsistency with the entanglement entropy between $\mathcal{H}_{\mathrm{CFT}}$ and $\mathcal{H}_{\mathrm{rad}}$ actually being the Bekenstein-Hawking entropy S_{BH}.

It is often suggested that state dependence makes quantum mechanics nonlinear. However the map from bulk states to boundary states is perfectly linear in this proposal; it just isn't an isometry. In effect, the naïve inner product on the bulk effective field theory is very different from the pull back of the boundary inner product to bulk states, which defines the actual microscopic inner product of the quantum theory. What then is the bulk inner product? The most natural answer, which is consistent with recent work on JT gravity [73], is that the bulk inner product is a statistical average of the boundary inner product over an ensemble of microscopic Hamiltonians, such as a small range of couplings.

It is well known that final state projection models can lead to issues with describing measurements done by an observer falling into the black hole [74–76]. While we leave a

detailed accounting of these issues to future work, the solution to at least some of these problems appears to be quantum error correction.

So long as we only consider a sufficiently small code spaces of allowed states, for example the code space of states describing the state of the observer jumping into the black hole, along with her experimental apparatus, there is always an isometry from the bulk to the global boundary (including the reservoir).

On the boundary, the evolution is described by standard quantum mechanics with no postselection. In the bulk, the observer is free to manipulate the unitary evolution of the experiment. The state of the experiment can also decohere and become entangled with the state of the observer; in more conventional phrasing, the observer can measure the experiment.[63] The isometry from the code space to global boundary states maps all these events to ordinary unitary quantum mechanics with no postselection on the global boundary. The only change that happens as the black hole evaporates is that the observer and the experiment end up being encoded in the reservoir $\mathcal{H}_{\text{rad}}$, rather than the CFT.

4.4 The peak of the Page curve

The other part of the evaporation that we skipped, in the interests of avoiding speculative discussion, was the period of time, very close to the Page time, when the Page curve peaks and begins to decline.

In a simple random unitary toy model of black hole evaporation, the entanglement entropy is almost exactly equal to the number of qubits in the radiation, until the black hole is within an $O(1)$ number of qubits of half its original size. There is then an $O(1)$ correction to the entropy at the peak of the curve, before the entropy becomes approximately equal to the number of qubits describing the black hole an $O(1)$ number of qubits later [60].

For an actual black hole, at times within $O(t_{\text{page}}/\sqrt{S_{BH}})$ of the Page time, $O(\sqrt{G_N})$ fluctuations in the horizon area of the black hole and $O(\sqrt{S})$ fluctuations in the total energy of the Hawking radiation mean that there will not be a single well-defined Ryu-Takayanagi surface. At such times, we should therefore expect that the entanglement entropy will neither increase as fast as the bulk entropy of the radiation, nor decrease as fast as the Bekenstein-Hawking entropy of the black hole.

However, we can write the total state of the black hole and Hawking radiation as a superposition of $O(\sqrt{S})$ states, each of which has only $O(G_N)$ fluctuation in the horizon area and where the entanglement spectrum of the Hawking radiation, in each state, has $O(1)$ width.[64]

For almost all of the states in such a superposition, there is still a well defined Ryu-Takayanagi surface. Indeed, if we believe that the lower bound on the reconstruction error from [38] is approximately saturated, then, at any given time, the quantum extremal

---

[63]It is important to note that causality prevents this decoherence from escaping the black hole. No measurement by an interior observer will ever 'have happened' from the perspective of an observer that remains outside the black hole.

[64]We cannot reduce the area fluctuations by more than this without substantially altering the bulk geometry, see [77, 78], for example, for details.

surface prescription will be valid with very small error for all but an $O(1)$ number of the states in this superposition.

For some fraction f of the states in the superposition, the interior will be encoded in the reservoir $\mathcal{H}_{\text{rad}}$, while for almost all the rest, it will be encoded in the CFT. As the black hole evaporates, the fraction f will increase, until eventually the fraction f approaches one and the interior can be reconstructed using only the reservoir, with only a very small error. From the fraction f, as a function of time, it should, in principle, be possible to calculate the shape of the peak of the Page curve.

Because the fluctuations in the evaporation rate smear out the Page time over an $O(\beta\sqrt{S_{B}H})$ time window, it seems plausible that, with sufficient work, one could calculate the entropy, up to an $O(1/\sqrt{S_{BH}})$ error, at all times. In other words, the fluctuations in the area of the black hole and the energy of the Hawking radiation may make it feasible to calculate the Page curve much more precisely than would otherwise be possible.

4.5 Explicit interior reconstruction

While we were able to make very precise statements in this paper about when and where information was encoded in an evaporating black hole, we said comparatively little about *how* the information was encoded. In particular, one would ideally want to have explicit, even if not necessarily practical, reconstructions of the interior operators.

There has been considerable work done in recent years on understanding how entanglement wedge reconstruction can be done explicitly [24, 79, 80]. However, these approaches often assume knowledge of some global reconstruction, such as HKLL [81]. For interior operators, it is not clear even what global reconstruction to use as a starting point. Infalling modes can, of course, simply be evolved back in time to give exterior operators, but this is not possible for outgoing interior modes.

This is not a new problem; it has been a major focus of research in AdS/CFT for many years. In particular, there are some credible suggestions of ways in which one can construct a state-dependent interior operator, given a particular choice of microstate [32, 33, 37]. If we believe the arguments from section 3.5, however, reconstructions should in principle be possible for much larger code spaces.

In appendix C, we give a simple generalisation of the Kourkoulou-Maldacena state-dependent interior reconstruction for the SYK model that works for code spaces with entropy almost as large as the Bekenstein-Hawking entropy.

We can also make the following more general argument for extending state-dependent reconstructions that work for a single microstate to minimally state-dependent reconstructions. Suppose we assume the existence of a single *unknown* reconstruction ϕ_{code} of an interior operator ϕ for a large code space $\mathcal{H}_{\text{code}}$ with orthonormal basis $|i\rangle$. Moreover, suppose we also assume that for any state $|i\rangle$ in this basis, we know an explicit reconstruction ϕ_i, which is consistent with the unknown reconstruction ϕ_{code}. In other words, for all $|i\rangle$,

$$\phi_{\text{code}} |i\rangle \approx \phi_i |i\rangle. \tag{4.10}$$

But then for any state

$$|\psi\rangle = \sum_i c_i |i\rangle$$

we have

$$\sum_i \phi_i |i\rangle \langle i|\psi\rangle = \sum_i c_i \phi_i |i\rangle \approx \phi_{\text{code}} |\psi\rangle. \tag{4.11}$$

Hence

$$\tilde{\phi}_{\text{code}} = \sum_i \phi_i |i\rangle \langle i| \tag{4.12}$$

is an explicit reconstruction that approximates ϕ_{code} when acting on any state in the code space.

4.6 The information paradox beyond AdS/CFT

This paper is entirely about AdS/CFT. However, by understanding the information paradox in AdS/CFT, we hope to eventually learn something about the information paradox in more general quantum gravity. Does information escape black holes in our universe (in the absence of the cosmic microwave background etc.) and if so how does it do so?

So far, entanglement wedge reconstruction, and the Ryu-Takayanagi formula, are only understood in the context of AdS/CFT. However, there is no obvious reason to think that they are specific to spacetimes with a negative cosmological constant. In particular, for asymptotically flat spacetimes, one can anchor a 'Cauchy' slice at some 'boundary' surface in asymptotic future infinity, and then calculate quantum extremal surfaces based on this. Indeed, in this case, one does not even need to do anything special to get absorbing boundary conditions. There is no timelike boundary for modes to reflect from. Instead, early modes will simply automatically not be included if they reach the asymptotic infinity at an earlier outgoing time than the time at which we anchored our 'Cauchy' slice.

For de Sitter spacetimes, which most resemble our universe, there is no timelike or lightlike asymptotic region that we can use to anchor spacelike slices. However, one would still hope that the basic conceptual ideas of this paper — essentially the fact that there is a state-dependent encoding of the black hole interior in the early Hawking radiation after the Page time — might be relevant.

As an intermediate step, consider the case of black holes in AdS/CFT that are small enough to be microcanonically unstable. These black holes are so small that we do not need to extract Hawking radiation into an auxiliary system for the black hole to evaporate; the black hole will have already evaporated by the time the Hawking radiation can reach the boundary and come back.

If we don't extract the Hawking radiation, there is no entanglement wedge that can show us that the interior is encoded in the Hawking radiation after the Page time. The Hawking radiation entirely surrounds the black hole horizon; there is no boundary region whose entanglement wedge contains the radiation, but not the black hole.

However, if we do extract the Hawking radiation, which we can do using non-local boundary dynamics, it is clear from entanglement wedge reconstruction that the interior is indeed encoded in the Hawking radiation, just like for larger AdS black holes. The interior

must still have been encoded in the Hawking radiation before we extracted the radiation; we just had no way to learn this using only entanglement wedge reconstruction.

To directly see that the interior was encoded in the radiation, even before we extracted the radiation, we would need a way to distinguish the microscopic degrees of freedom encoding a small neighbourhood of the black hole from the microscopic degrees of freedom describing the Hawking radiation further out. This would require understanding holography beyond asymptotic boundaries. There has been considerable recent progress in that direction using $T\bar{T}$ deformations of conformal field theories [82–86].

Acknowledgments

A Cut-offs in Rindler space

In this appendix, we study a simple pedagogical example of a situation where understanding the coordinate dependence of cut-offs is vital, if we want to calculate entanglement entropies correctly. The example is closely related to, but distinct from, the black hole extremal surface calculations in sections 2.2 and 2.4.

Consider the interval $[0, r]$ of the vacuum state in some $(1 + 1)$-dimensional conformal field theory. The entanglement entropy of this interval is given by

$$S = \frac{c}{3} \log \frac{r}{\sqrt{\varepsilon_1, \varepsilon_2}} \tag{A.1}$$

where ε_1 and ε_2 are the cut-offs at each end of the interval [57, 58]. We therefore find that the derivative of the entanglement entropy

$$\frac{dS}{dr} = \frac{c}{3r}. \tag{A.2}$$

We can also do this calculation in Rindler space, where it corresponds to finding the derivative of the entropy of an infinite half-line of a thermal state at inverse temperature $\beta = 2\pi$. The entropy of a long interval of a thermal state of a CFT is equal to

$$S = \frac{\pi c\, l}{3\beta} \tag{A.3}$$

where l is the length of the interval [57, 58]. We therefore find that

$$\frac{dS}{dr^*} = \frac{c}{6} \tag{A.4}$$

where the Rindler position $r^* = \log r$. Hence

$$\frac{dS}{dr} = \frac{1}{r}\frac{dS}{dr^*} = \frac{c}{6r}. \tag{A.5}$$

However, this differs from (A.2) by a factor of two. We apparently have a contradiction.

We can also look at the entanglement entropy of the interval $[r, \infty]$. In the Minkowski space calculation, the gradient of the entropy of an interval tends to zero as the length of the interval tends to infinity. However, for the thermal state in Rindler space, the gradient is simply the negative of (A.5).

The resolution of the paradox is, of course, the cut-offs. Implicitly, we assumed in the Minkowski space calculation that the cut-off was constant in units of r, while in the Rindler space calculation we assumed that it was constant in terms of the Rindler position r^*. However, r and r^* are nonlinearly related. So the cut-off cannot be constant in both units.

Let us assume that we actually wanted the cut-off to be constant in terms of the Rindler position r^*. Recall that a constant cut-off in units of r^* really means that the cut-off is equal to

$$\varepsilon_0 \frac{\partial}{\partial r^*},$$

for some constant ε_0. Since

$$\frac{\partial}{\partial r^*} = r\frac{\partial}{\partial r} \tag{A.6}$$

the cut-off ε_2 in (A.1), which is in units of r, is $r\,\varepsilon_0$. We therefore find that

$$\frac{dS}{dr} = \frac{c}{6r}, \tag{A.7}$$

while the derivative of the entropy for the interval $[r, \infty]$ is $-c/6r$. The results now agree with the Rindler space calculation.

B Finite temperature infalling modes

In this appendix, we generalise the explicit calculation of the location of the non-empty quantum extremal surface from section 2.2 to thermal infalling modes at finite temperature, and to pure infalling modes with constant energy density and without long range entanglement. As in section 2.2, we assume that the outgoing modes are extracted from close to the horizon and so there are no greybody factors. We assume throughout the section that we are after the Page time, and so the non-empty quantum extremal surface is the Ryu-Takayanagi surface.[65] Throughout this section, we shall work in Eddington-Finkelstein coordinates, as in section 2.2. However, the calculations can also easily be done in Kruskal-like coordinates, as in section 2.4.

---

[65]The exception is when the temperature, or energy density of the infalling modes is sufficient to prevent the black hole ever reaching the Page time. As we shall see, in those cases, there does not exist a non-empty quantum extremal surface at all.

We begin by studying the case of thermal infalling modes at a temperature T' that may be different from the black hole temperature T. We assume that the infalling modes are purified by an auxiliary Hilbert space $\mathcal{H}_{\text{pur}}$ and hence, importantly, are unentangled with the outgoing Hawking radiation. We consider both the information learned by an observer with access to $\mathcal{H}_{\text{rad}} \otimes \mathcal{H}_{\text{pur}}$, and an observer with access only to $\mathcal{H}_{\text{rad}}$.

Taking into account the new infalling thermal flux, we find that (2.15) becomes

$$\frac{dM}{dv} = \frac{c_{\text{evap}}\pi}{12}(T'^2 - T^2). \tag{B.1}$$

Hence we have

$$\frac{dr_s}{dv} = \frac{c_{\text{evap}}\pi G_N}{3T(d-1)r_s^{d-2}\Omega_{d-1}}(T'^2 - T^2), \tag{B.2}$$

and, using (2.11), the event horizon is at

$$r_{\text{hor}} = r_s + \frac{c_{\text{evap}}G_N}{6(d-1)r_s^{d-2}\Omega_{d-1}}\frac{T'^2 - T^2}{T^2}. \tag{B.3}$$

Using (2.13), the classical maximin surface lies on the classical apparent horizon r_s at

$$v = -\frac{\beta}{2\pi}\log\frac{S_{BH}T^2}{c_{\text{evap}}(T^2 - T'^2)} + O(\beta). \tag{B.4}$$

As the infalling radiation temperature T' approaches the black hole temperature T, the location of the classical maximin surface diverges into the infinite past because $dr_s/dv \to 0$. In contrast, we shall see that the non-empty quantum extremal surface for the CFT remains well-behaved at this temperature.

We first calculate the Ryu-Takayanagi surface associated to $\mathcal{H}_{\text{CFT}}$. (Since the overall tripartite state is pure, this is also the Ryu-Takayanagi surface for $\mathcal{H}_{\text{rad}} \otimes \mathcal{H}_{\text{pur}}$.) The entropy of the outgoing modes is the same as (2.26), but, because the ingoing modes are now at finite temperature, the entropy of the infalling modes in the entanglement wedge of the CFT is now

$$S_{\text{in}} = -\frac{c_{\text{evap}}\pi T'v}{6} + \ldots, \tag{B.5}$$

where, as usual, we have dropped constant terms and we have assumed that the cut-off is independent of position in units of v.[66] The total bulk entropy is therefore

$$S_{\text{bulk}} = \frac{c_{\text{evap}}}{6}\log(r_{lc}(v) - r) - \frac{c\pi v}{6}(T + T') + \ldots \tag{B.6}$$

Note that for $T' = T$, which corresponds to the Hartle-Hawking state, we expect that the total entanglement entropy of ingoing and outgoing modes will agree with the Minkowski vacuum formula for the entanglement entropy based on the proper distance between the quantum extremal surface and the point where the outgoing modes are extracted. Using the Schwarzschild time translation symmetry, or boost symmetry, of the Hartle-Hawking state, we can map this interval to a small interval close to the bifurcation

---

[66]We are also, as usual, ignoring the corrections associated with the finite infalling time range of the infalling modes because these corrections should vanish in the semiclassical limit.

surface, where the Hartle-Hawking state locally looks like the Minkowski vacuum.[67] It can easily be verified that this is indeed the case.

Using (B.2) and (B.6), it is easy to calculate the location of the Ryu-Takayanagi surface of $\mathcal{H}_{\mathrm{CFT}}$. We find that

$$0 = \frac{\partial S_{\mathrm{bulk}}}{\partial v} + \frac{1}{4G_N}\frac{\partial A}{\partial v}, \tag{B.7}$$

$$= \frac{\partial S_{\mathrm{bulk}}}{\partial v}, \tag{B.8}$$

$$= \frac{\partial r_{lc}/\partial v}{6(r_{lc} - r)} - \frac{\pi}{6}(T + T'), \tag{B.9}$$

$$(T + T')(r_{lc}(v) - r) = 2T(r_{lc} - r_s(v)), \tag{B.10}$$

while (2.34) continues to be valid. Hence the quantum extremal surface lies at

$$r = r_s + \frac{T' - T}{T}\frac{G_N c_{\mathrm{evap}}}{3(d-1)\Omega_{d-1}r_s^{d-2}} = r_{\mathrm{hor}} - (T' - T)^2\frac{G_N c_{\mathrm{evap}}}{6(d-1)r_s^{d-2}\Omega_{d-1}T^2}, \tag{B.11}$$

and satisfies

$$r_{lc}(v) = r_s + \frac{T' + T}{T}\frac{G_N c}{3(d-1)\Omega_{d-1}r_s^{d-2}} = r_{\mathrm{hor}} + \left[4 - \frac{(T'-T)^2}{T^2}\right]\frac{G_N c_{\mathrm{evap}}}{6(d-1)r_s^{d-2}\Omega_{d-1}}. \tag{B.12}$$

Thus

$$v = -\frac{\beta}{2\pi}\log\frac{S_{BH}}{c_{\mathrm{evap}}(4 - \frac{(T'-T)^2}{T^2})} + O(\beta). \tag{B.13}$$

When the infalling modes are at the same temperature as the black hole, the Ryu-Takayanagi surface lies exactly on the event horizon.[68] At all other temperatures, it lies strictly inside the event horizon. When $T' = T$, and only when $T' = T$, the total thermodynamic entropy of the black hole and exterior radiation does not increase with time. The entropy of the new Hawking radiation is cancelled by the loss of entropy from radiation falling into the black hole, while the entropy of the black hole itself stays constant. Because all outgoing modes in the interior are in the entanglement wedge of $\mathcal{H}_{\mathrm{rad}} \otimes \mathcal{H}_{\mathrm{pur}}$, the Hawking radiation, even Hawking radiation far in the future, is perfectly thermally entangled with $\mathcal{H}_{\mathrm{rad}} \otimes \mathcal{H}_{\mathrm{pur}}$. Since the ingoing modes are also thermally entangled with $\mathcal{H}_{\mathrm{pur}}$ and the horizon area is constant, the black hole entanglement entropy stays constant.

---

[67]In contrast, the case where the infalling modes have zero temperature corresponds to the Unruh state, which is singular at the white hole horizon and so does not locally look like the Minkowski vacuum near the bifurcation surface.

[68]Since we only calculated the radius to $O(G_N)$, higher order corrections can potentially move the quantum extremal surface outside the horizon. When calculating the entanglement wedge of the CFT *plus* all the future ingoing thermal modes that will be thrown into the black hole, the RT surface cannot end up outside the event horizon without creating a paradox. However this is not true for the entanglement wedge of the CFT alone. In that case, corrections from only having a finite interval of infalling thermal modes pushes the RT surface an $O(G_N^2)$ radial distance outside the horizon, as was shown in [87] after this paper first appeared on arXiv. (The result follows most obviously as a consequence of the time-reversal symmetry of the state forcing the RT surface to lie in the static slice.)

This is consistent with the Page curve, which can also be derived using Ryu-Takayanagi formula. Indeed, at any temperature T', the entanglement structure of the Hawking radiation will be exactly consistent with the Page curve, because of our general argument from section 2.3. It can easily be verified that this is indeed the case.

The Ryu-Takayanagi surface remains approximately one scrambling time in the past so long as the temperature of the infalling modes is relatively low (the latest infalling time is obtained at $T' = T$), but diverges into the past at $T' = 3T$. At higher temperatures, the radial distance between the quantum extremal surface and the event horizon, required by (B.11), will be greater than the radial distance between the quantum extremal surface and the outgoing lightcone, required by (B.12). Since the outgoing lightcone can never be inside the event horizon, no quantum extremal surface can exist.

An observer with access to $\mathcal{H}_{\rm rad} \otimes \mathcal{H}_{\rm pur}$ will only ever learn the state of a diary thrown into the black hole if the temperature $T' < 3T$. Interestingly, $T' = 3T$ is exactly the temperature at which thermal Hawking radiation, unentangled with $\mathcal{H}_{\rm rad} \otimes \mathcal{H}_{\rm pur}$, becomes consistent with unitarity. At this temperature,

$$\frac{1}{4G_N}\frac{dA_{\rm hor}}{dv} = \frac{2c_{\rm evap}\pi}{3\beta} = \frac{dS_{\rm in}}{dv} + \frac{dS_{\rm rad}}{dv}, \tag{B.14}$$

where $dS_{\rm in}/dv \geq 0$ is the entropy of the ingoing modes per unit infalling time and $dS_{\rm rad}/dv \geq 0$ is the rate that entropy is produced in the Hawking radiation. The increase in the Bekenstein-Hawking entropy is therefore just sufficient to purify both the outgoing Hawking radiation, and the purification $\mathcal{H}_{\rm pur}$ of the infalling modes.

If the observer only has access to $\mathcal{H}_{\rm rad}$, but not to $\mathcal{H}_{\rm pur}$, it will affect the information that they learn about the black hole. To understand this, we need to calculate the location of the Ryu-Takayanagi surface for $\mathcal{H}_{\rm rad}$. (Because the system is no longer in a bipartite pure state, this is not the same as the Ryu-Takayanagi surface of $\mathcal{H}_{\rm CFT}$.)

The entanglement wedge of $\mathcal{H}_{\rm rad}$ contains the part of the interior inside the Ryu-Takayanagi surface, as well as the outgoing modes that were extracted into the reservoir. Since the overall state of the outgoing modes is pure, the outgoing entropy in the entanglement wedge of $\mathcal{H}_{\rm rad}$ for a given candidate RT surface is equal to the outgoing entropy in the entanglement wedge of $\mathcal{H}_{\rm CFT}$ for the same candidate surface. (Since the Ryu-Takayanagi surfaces for $\mathcal{H}_{\rm rad}$ and $\mathcal{H}_{\rm CFT}$ will end up being different, they will have different entropies for the outgoing modes. However as a *function* of the location of the surface, they are the same.)

This is not true for the ingoing modes, which are in a mixed state that is purified by $\mathcal{H}_{\rm pur}$. The entanglement wedge of $\mathcal{H}_{\rm CFT}$ contains ingoing modes at infalling times that are later than the Ryu-Takayanagi surface, while the entanglement wedge of $\mathcal{H}_{\rm rad}$ contains at infalling times that are earlier than the RT surface. Hence, instead of (B.6), we have

$$S_{\rm bulk} = \frac{c_{\rm evap}}{6}\log\left(r_{lc}(v) - r\right) + \frac{c\pi v}{6}(T' - T) + \dots \tag{B.15}$$

We therefore find that the quantum extremal surface for $\mathcal{H}_{\rm rad}$ lies at

$$r = r_s - \frac{T' + T}{T}\frac{G_N c_{\rm evap}}{3(d-1)\Omega_{d-1}r_s^{d-2}} = r_{\rm hor} - (T' + T)^2\frac{G_N c}{6(d-1)r_s^{d-2}\Omega_{d-1}T^2}, \tag{B.16}$$

and satisfies

$$r_{lc}(v) = r_s + \frac{T - T'}{T} \frac{G_N c}{3(d-1)\Omega_{d-1} r_s^{d-2}} = r_{\text{hor}} + \left[4 - \frac{(T'+T)^2}{T^2}\right] \frac{G_N c_{\text{evap}}}{6(d-1) r_s^{d-2} \Omega_{d-1}}. \quad \text{(B.17)}$$

Thus

$$v = -\frac{\beta}{2\pi} \log \frac{S_{BH}}{c(4 - \frac{(T'+T)^2}{T^2})} + O(\beta). \quad \text{(B.18)}$$

For $T' < T$, information thrown into the black hole will still reappear in the Hawking radiation. However the location of the extremal surface diverges as $T' \to T$; for $T' \geq T$ no information will ever escape in the Hawking radiation for an observer with access only to $\mathcal{H}_{\text{rad}}$.

As before, this is exactly the temperature at which the Hawking radiation can look thermal, to an observer with access only to $\mathcal{H}_{\text{rad}}$ without violating unitarity. The total entropy of $\mathcal{H}_{\text{CFT}} \otimes \mathcal{H}_{\text{rad}}$ increases because thermal modes are being thrown into the black hole. The entropy of $\mathcal{H}_{\text{rad}}$ can therefore increase at the same rate, and the new Hawking radiation can be unentangled with $\mathcal{H}_{\text{rad}}$, even while the horizon area, and hence the entropy, of the black hole remains constant.

Finally, the Ryu-Takayanagi surface of $\mathcal{H}_{\text{pur}}$ will always be empty — it will always be in a thermal state. This is a necessary consequence of the boundary dynamics being unitary; the state of $\mathcal{H}_{\text{pur}}$ is initially thermal, and this is unchanged when we throw its purification into a black hole.

We can also calculate the location of the quantum extremal surface when the infalling modes are in a pure state, with constant energy density η, but without long range entanglement, for example, if there is a constant infalling particle flux. Again, it is important that the ingoing modes are unentangled with the outgoing Hawking radiation. In this case, both (2.31) and (2.34) will continue to be valid as in the vacuum case. However, instead of (2.15), we now have

$$\frac{dM}{dv} = -\frac{c_{\text{evap}}\pi}{12\beta^2} + \eta, \quad \text{(B.19)}$$

and

$$\frac{dr_s}{dv} = -\frac{c_{\text{evap}}\pi G_N}{3\beta(d-1)r_s^{d-2}\Omega_{d-1}} + \frac{4G_N \beta \eta}{(d-1)r_s^{d-2}\Omega_{d-1}}. \quad \text{(B.20)}$$

This affects the radius $r_{l.c.}(v)$ of the past lightcone as a function of the infalling time v, which is given in (2.12). Hence, while the Ryu-Takayanagi surface still lies at

$$r = r_s - \frac{G_N c_{\text{evap}}}{3(d-1)\Omega_{d-1} r_s^{d-2}}, \quad \text{(B.21)}$$

its infalling time will now be given by

$$v = -\frac{\beta}{2\pi} \log \frac{S_{BH}}{1 - 4\beta^2 \eta / c_{\text{evap}}\pi} + O(\beta). \quad \text{(B.22)}$$

As before, when sufficient energy, in this case $\eta > c_{\text{evap}}\pi / 4\beta^2$ are thrown into the black hole, no quantum extremal surface exists. The event horizon, and thus the outgoing light cone, are at too large a radius for (2.31) and (2.34) to be simultaneously satisfied.

Yet again, this is exactly the point at which it stops being necessary for information to escape the black hole in order to preserve unitarity. At this energy density the rate of increase of the Bekenstein Hawking entropy is equal to the rate of increase in the entropy of the radiation

$$\frac{1}{4G_N}\frac{dA_{\text{hor}}}{dv} = \frac{c_{\text{evap}}\pi}{6\beta} = \frac{dS_{\text{rad}}}{dv}. \tag{B.23}$$

Hence the Hawking radiation can remain thermal, and unentangled with $\mathcal{H}_{\text{rad}}$, forever, without exceeding the entanglement entropy exceeding the Bekenstein-Hawking entropy of the black hole.

The ingoing energy flux at which information stops escaping the black hole is highest for thermal infalling modes and an observer who has access to both the reservoir $\mathcal{H}_{\text{rad}}$ and the purification $\mathcal{H}_{\text{pur}}$ of the ingoing modes. This is because there needs to be sufficient Bekenstein-Hawking entropy in the black hole both to purify outgoing thermal Hawking radiation *and* to purify the degrees of freedom in $\mathcal{H}_{\text{pur}}$.

In contrast, when the observer only has acccess to $\mathcal{H}_{\text{rad}}$, the required ingoing energy flux for thermal infalling modes is much smaller. The ingoing entropy makes it *easier* for the Hawking radiation to be unentangled with $\mathcal{H}_{\text{rad}}$, because $\mathcal{H}_{\text{rad}}$ can be purified by $\mathcal{H}_{\text{pur}}$ as well as the black hole.

Finally, when the ingoing modes are in a pure state with no long range entanglement, information stops escaping at an intermediate ingoing energy flux. The increase in the Bekenstein-Hawking entropy needs to be sufficient to purify the Hawking radiation in $\mathcal{H}_{\text{rad}}$; there is no ingoing bulk entropy to make this either harder or easier.

C   Minimal state dependence in the SYK model

In this appendix, we construct minimally state-dependent interior reconstructions in a simple toy model of quantum gravity, known as the SYK model. This model has been studied in great depth in the last few years [88–94]; here we provide only the bare minimum of background detail necessary for our purposes.

The SYK model is a $0+1$-dimensional quantum mechanical model that consists of N Majorana fermions ψ_i. These satisfy

$$\{\psi_i, \psi_j\} = \delta_{i,j}.$$

Using a Jordan-Wigner transformation, it can be easily seen that there is a single qubit degree of freedom associated with every pair of Majorana fermions. The Hilbert space therefore has dimension $2^{N/2}$. The model has Hamiltonian

$$H = \sum_{iklm} j_{iklm}\psi_i\psi_k\psi_l\psi_m, \tag{C.1}$$

where j_{iklm} are independent Gaussian random couplings with $\langle j_{iklm}^2\rangle = 6J^2/N^3$.

In the limit $N \to \infty$, at fixed $\beta J \gg 1$, the SYK model appears to become holographic; it has many features that resemble nearly-AdS$_2$ gravity. Although the precise dual gravity description, if one exists, remain unknown, both the SYK model and simple nearly-AdS$_2$

gravity theories such as Jackiw-Teitelboim gravity [73, 95–98] have an emergent reparameterisation symmetry that is both spontaneously and explicitly broken, with the explicit symmetry-breaking term proportional to the so-called Schwarzian action

$$S = \frac{\alpha_S N}{J} \int d\tau \left(\frac{f''}{f'}\right)' - \frac{1}{2}\left(\frac{f''}{f'}\right)^2, \tag{C.2}$$

where $f(\tau)$ is the reparameterisation and α_S is a numerical constant. From a gravity perspective, this action appears as a boundary term when we cut-off the nearly-AdS$_2$ geometry at some fixed dilaton value ϕ_b; for Jackiw-Teitelboim gravity, in particular, it describes the entire dynamics of the theory, which can be interpreted as the dynamics of a boundary particle, describing the location of the cut-off, in a rigid AdS$_2$ background.

A complete basis for the entire Hilbert space of the SYK model is given by the states $|B_s\rangle$, satisfying

$$(\psi_{2k-1} - is_k\psi_{2k})|B_s\rangle = 0, \tag{C.3}$$

where for all k, we have $s_k = \pm 1$. If we evolve these states in Euclidean time, we get a set of states

$$|B_s(\beta)\rangle = e^{-\beta H/2}|B_s\rangle, \tag{C.4}$$

which form an approximate, overcomplete basis for the low energy states of the theory. In fact, if we allow arbitrary superpositions of these states, they still form a complete basis for the entire Hilbert space, because the map $e^{-\beta H}$ is invertible. However, to create a high energy state, we need a very finely tuned superposition

$$|\psi\rangle = \sum_s c_s |B_s(\beta)\rangle, \tag{C.5}$$

where

$$\langle\psi|\psi\rangle \ll \sum_s |c_s|^2 \langle B_s(\beta)|B_s(\beta)\rangle. \tag{C.6}$$

However we will only allow 'generic' superpositions of the states $|B_s(\beta)\rangle$ that do not satisfy (C.6).

It was shown in [37] that the states $|B_s(\beta)\rangle$ have a natural gravity interpretation as black hole microstates with a smooth interior ending on an 'end-of-the-world brane' (figure 17). Excitations in the interior can be created by acting with additional boundary operators during the Euclidean time evolution.

If the system is evolved with the unperturbed Hamiltonian H, then such excitations can never reach the boundary. However, if we perturb the Hamiltonian by

$$\delta H = -\varepsilon J \sum_{k=1}^{N/2} s_k i\psi_{2k-1}\psi_{2k}, \tag{C.7}$$

then, from a gravity perspective, the Schwarzian 'boundary particle' is pulled into the bulk of the AdS$_2'$ space, as shown in figure 17. By evolving the system with the perturbed Hamiltonian $H + \delta H$, we can render the interior of the black hole visible to the boundary.
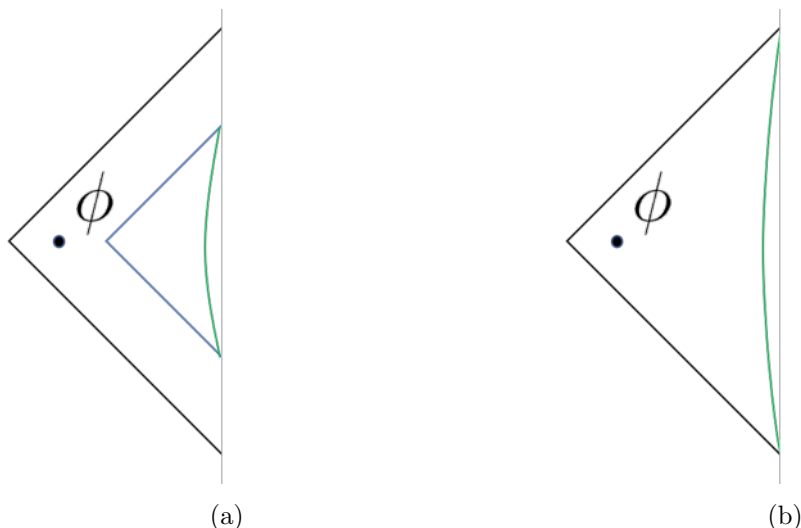
Figure 17. The states $|B_s(\beta)\rangle$ have a gravity description as a one sided black hole, ending on an end-of-the-world brane (black). If the system is evolved using the unperturbed SYK Hamiltonian, shown in figure 17a, the bulk operator ϕ lies behind the black hole horizon (blue). However if the Hamiltonian is perturbed in a state-dependent way, shown in figure 17b, the boundary (green) is pulled inwards and so the operator no longer lies behind a horizon. In the original construction, the Hamiltonian was precisely tuned for a single state $|B_s(\beta)\rangle$. However one can easily adapt the Hamiltonian to work for a large set of states.

Interior operators can be reconstructed using boundary operators time-evolved using this perturbed Hamiltonian.

The perturbation δH was carefully adapted to the state $|B_s(\beta)\rangle$. The reconstruction is therefore highly state-dependent. A natural question is whether we can reduce this state dependence, and find a reconstruction that works for a larger class of microstates.

Instead of using the perturbation (C.7), suppose instead that we perturb the Hamiltonian H by

$$\delta H_f = -\varepsilon J \sum_{k=1}^{fN/2} s_k i\psi_{2k-1}\psi_{2k}, \tag{C.8}$$

where $0 < f < 1$ is a fixed $O(1)$ fraction. Since the number of terms in δH_f continues to be $O(N)$, the perturbation δH_f will also make the interior of the black hole microstate $|B_s(\beta)\rangle$ visible to the boundary.[69]

However this perturbation only depended on the first $fN/2$ spins s_k of the microstate $|B_s(\beta)\rangle$. We have found a single reconstruction that works for $2^{(1-f)N/2}$ different mi-

---

[69]The argument that the perturbation δH can be used to make the interior visible to a boundary observer is given in section 7 of [37]. For the full details, we simply refer readers to that work. However the basic strategy is to assume that for $\varepsilon \ll 1$ we can treat δH as a perturbation to the Schwarzian action (C.2) of the reparameterisation modes. So long as we have $O(N)$ terms, the perturbation will appear in the semiclassical action at large N. It can then be shown that, at sufficiently large, but fixed, βJ, we can choose ε so that the semiclassical large N equation of motion for the Schwarzian 'boundary particle' makes the entire black hole interior visible. The argument for δH_f is identical, except for the addition of the $O(1)$ factor f in the perturbation to the large N semiclassical Schwarzian action.

crostates $|B_s(\beta)\rangle$. By linearity, the same reconstruction should also work for generic superpositions of these microstates.

Since the low temperature entropy of the SYK model is approximately

$$S_0 \approx 0.23N, \tag{C.9}$$

we have found a single reconstruction that is valid for more than e^{S_0} microstates. Of course, not all these microstates are independent. Instead, the effective size of the code subspace is determined by the entropy of the mixed state

$$\rho_{s_1} = 2^{-(1-f)N/2} \sum_{s_2} |B_{s_1 \oplus s_2}(\beta)\rangle \langle B_{s_1 \oplus s_2}(\beta)| , \tag{C.10}$$

where the sum is over spins $s_k \in s_2$ for $fN/2 \leq k \leq N/2$ while the spins $s_k \in s_1$ for $1 \leq k < fN/2$ are held fixed.

How large is this entropy? Since

$$\frac{e^{-\beta H}}{\text{Tr}(e^{-\beta H})} = 2^{-fN/2} \sum_{s_1} \rho_{s_1}, \tag{C.11}$$

then the strict concavity of entropy means that

$$\langle S(\rho_{s_1})\rangle_{s_1} < S_0, \tag{C.12}$$

where the expectation is taken over possible states s_1 of the fixed spins. If ρ_{s_1} were a uniform mixture of $2^{(1-f)N/2}$ randomly chosen microstates $|B_s(\beta)\rangle$, we would expect that $S(\rho_{s-1})$ would be very close to S_0 for $(1-f)N/2 > S_0$. There would be no remaining space to encode the interior degrees of freedom. However this will not be the case for the particular set of microstates in (C.10).

At large N, there is an emergent $O(N)$ symmetry of the SYK model. In particular there is a $\mathbb{Z}_2^n$ subgroup of this symmetry group, called the flip subgroup, that acts transitively on the set of states $|B_s(\beta)\rangle)$. This means that $S(\rho_{s_1})$ depends, at leading order, only on the number of spins that are held fixed, and not on the signs of those spins.

If no spins are held fixed, the ensemble is simply the canonical ensemble, which has entropy S_0 to leading order in $1/N$ for fixed $\beta J \gg 1$. Now suppose that we know the average entropy S_f for a state ρ_{s_1} with a fixed fraction f of the spins held fixed. We then consider the ensembles $\rho_{s_1 \oplus 1}$ and $\rho_{s_1 \oplus -1}$ formed by fixing a single additional spin. These two ensembles therefore have a fraction $f + 2/N$ of their spins fixed. Note that

$$\rho_{s_1} = \frac{1}{2} \left(\rho_{s_1 \oplus 1} + \rho_{s_1 \oplus -1} \right). \tag{C.13}$$

Hence

$$2S(\rho_{s_1}) - S(\rho_{s_1 \oplus 1}) - S(\rho_{s_1 \oplus -1}) = S(\rho_{s_1 \oplus 1}||\rho_{s_1}) + S(\rho_{s_1 \oplus -1}||\rho_{s_1}) \tag{C.14}$$

$$\geq \frac{1}{4}||\rho_{s_1 \oplus 1} - \rho_{s_1 \oplus -1}||_1^2, \tag{C.15}$$

where we have used Pinsker's inequality [99]. However, we can lower bound the trace distance $\|\rho_{s_1\oplus 1} - \rho_{s_1\oplus -1}\|_1$ by

$$\|\rho_{s_1\oplus 1} - \rho_{s_1\oplus -1}\|_1 = \max_{\mathcal{O}} \frac{|\mathrm{Tr}\left[\mathcal{O}\left(\rho_{s_1\oplus 1} - \rho_{s_1\oplus -1}\right)\right]|}{\|\mathcal{O}\|} \geq |\mathrm{Tr}\left[\psi_{2k-1}\psi_{2k}\left(\rho_{s_1\oplus 1} - \rho_{s_1\oplus -1}\right)\right]|,$$
(C.16)

where $k = Nf/2 + 1$ labels the additional spin fixed in $\rho_{s_1\oplus\pm 1}$, but not in ρ_{s_1}. This last quantity was shown to in [37] to be order one in the limit of large N (although it decays as a function of βJ). Hence

$$S_f - S_{f+2/N} = O(1)$$
(C.17)

and thus

$$S_0 - S_f = O(fN),$$
(C.18)

for any fixed $f > 0$. If the fraction f is small, the entropy is very close to S_0 at leading order, but there is still plenty of space left to encode the interior. We have found an explicit, minimally state-dependent reconstruction.

Of course, we have only constructed minimally state-dependent operators for a very particular class of ensembles of microstates. Our arguments in section 3.5 suggested that there should exist minimally state-dependent reconstructions for *any* sufficiently small code subspace. For the special ensembles that we have considered in this section, the reconstructions only involved a simple perturbation to the SYK Hamiltonian; for arbitrary code subspaces, the reconstructions would presumably be much more complicated. We gave a more general, but much less practical, procedure for constructing minimally state-dependent reconstructions out of reconstructions that only work for individual states in section 4.5; it seems reasonable to expect that such a procedure should also work for the SYK model.

Open Access. This article is distributed under the terms of the Creative Commons Attribution License (CC-BY 4.0), which permits any use, distribution and reproduction in any medium, provided the original author(s) and source are credited.

References

[1] J.M. Maldacena, *The large N limit of superconformal field theories and supergravity*, *Int. J. Theor. Phys.* **38** (1999) 1113 [hep-th/9711200] [INSPIRE].

[2] E. Witten, *Anti-de Sitter space and holography*, *Adv. Theor. Math. Phys.* **2** (1998) 253 [hep-th/9802150] [INSPIRE].

[3] W.G. Unruh and R.M. Wald, *Information loss*, *Rept. Prog. Phys.* **80** (2017) 092002 [arXiv:1703.02140] [INSPIRE].

[4] S.W. Hawking, *Black hole explosions*, *Nature* **248** (1974) 30 [INSPIRE].

[5] S.W. Hawking, *Particle creation by black holes*, *Commun. Math. Phys.* **43** (1975) 199 [*Erratum ibid.* **46** (1976) 206] [INSPIRE].

[6] L. Susskind, L. Thorlacius and J. Uglum, *The stretched horizon and black hole complementarity*, *Phys. Rev. D* **48** (1993) 3743 [hep-th/9306069] [INSPIRE].

[7] D.N. Page, *Time dependence of Hawking radiation entropy*, *JCAP* **09** (2013) 028 [arXiv:1301.4995] [INSPIRE].

[8] P. Hayden and J. Preskill, *Black holes as mirrors: quantum information in random subsystems*, *JHEP* **09** (2007) 120 [arXiv:0708.4025] [INSPIRE].

[9] D.N. Page, *Information in black hole radiation*, *Phys. Rev. Lett.* **71** (1993) 3743 [hep-th/9306083] [INSPIRE].

[10] A. Almheiri, D. Marolf, J. Polchinski and J. Sully, *Black holes: complementarity or firewalls?*, *JHEP* **02** (2013) 062 [arXiv:1207.3123] [INSPIRE].

[11] E.H. Lieb and M.B. Ruskai, *Proof of the strong subadditivity of quantum-mechanical entropy*, in *Les rencontres physiciens-mathématiciens de Strasbourg — RCP25* **19**, (1973), pg. 36.

[12] Y. Nomura, J. Varela and S.J. Weinberg, *Complementarity endures: no firewall for an infalling observer*, *JHEP* **03** (2013) 059 [arXiv:1207.6626] [INSPIRE].

[13] R. Bousso, *Complementarity is not enough*, *Phys. Rev. D* **87** (2013) 124023 [arXiv:1207.5192] [INSPIRE].

[14] L. Susskind, *Singularities, firewalls and complementarity*, arXiv:1208.3445 [INSPIRE].

[15] E. Verlinde and H. Verlinde, *Black hole entanglement and quantum error correction*, *JHEP* **10** (2013) 107 [arXiv:1211.6913] [INSPIRE].

[16] L. Susskind, *The transfer of entanglement: the case for firewalls*, arXiv:1210.2098 [INSPIRE].

[17] D. Harlow and P. Hayden, *Quantum computation vs. firewalls*, *JHEP* **06** (2013) 085 [arXiv:1301.4504] [INSPIRE].

[18] J. Maldacena and L. Susskind, *Cool horizons for entangled black holes*, *Fortsch. Phys.* **61** (2013) 781 [arXiv:1306.0533] [INSPIRE].

[19] 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] [INSPIRE].

[20] M. Headrick, V.E. Hubeny, A. Lawrence and M. Rangamani, *Causality & holographic entanglement entropy*, *JHEP* **12** (2014) 162 [arXiv:1408.6300] [INSPIRE].

[21] A.C. Wall, *Maximin surfaces, and the strong subadditivity of the covariant holographic entanglement entropy*, *Class. Quant. Grav.* **31** (2014) 225007 [arXiv:1211.3494] [INSPIRE].

[22] D.L. Jafferis, A. Lewkowycz, J. Maldacena and S.J. Suh, *Relative entropy equals bulk relative entropy*, *JHEP* **06** (2016) 004 [arXiv:1512.06431] [INSPIRE].

[23] 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] [INSPIRE].

[24] 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].

[25] A. Almheiri, X. Dong and D. Harlow, *Bulk locality and quantum error correction in AdS/CFT*, *JHEP* **04** (2015) 163 [arXiv:1411.7041] [INSPIRE].

[26] S. Ryu and T. Takayanagi, *Holographic derivation of entanglement entropy from AdS/CFT*, *Phys. Rev. Lett.* **96** (2006) 181602 [hep-th/0603001] [INSPIRE].

[27] M. Headrick and T. Takayanagi, *A holographic proof of the strong subadditivity of entanglement entropy*, *Phys. Rev. D* **76** (2007) 106013 [arXiv:0704.3719] [INSPIRE].

[28] V.E. Hubeny, M. Rangamani and T. Takayanagi, *A covariant holographic entanglement entropy proposal*, *JHEP* **07** (2007) 062 [arXiv:0705.0016] [INSPIRE].

[29] N. Engelhardt and A.C. Wall, *Quantum extremal surfaces: holographic entanglement entropy beyond the classical regime*, *JHEP* **01** (2015) 073 [arXiv:1408.3203] [INSPIRE].

[30] X. Dong and A. Lewkowycz, *Entropy, extremality, Euclidean variations, and the equations of motion*, *JHEP* **01** (2018) 081 [arXiv:1705.08453] [INSPIRE].

[31] A. Almheiri, *Holographic quantum error correction and the projected black hole interior*, arXiv:1810.02055 [INSPIRE].

[32] K. Papadodimas and S. Raju, *An infalling observer in AdS/CFT*, *JHEP* **10** (2013) 212 [arXiv:1211.6767] [INSPIRE].

[33] K. Papadodimas and S. Raju, *State-dependent bulk-boundary maps and black hole complementarity*, *Phys. Rev. D* **89** (2014) 086010 [arXiv:1310.6335] [INSPIRE].

[34] D. Harlow, *Aspects of the Papadodimas-Raju proposal for the black hole interior*, *JHEP* **11** (2014) 055 [arXiv:1405.1995] [INSPIRE].

[35] K. Papadodimas and S. Raju, *Remarks on the necessity and implications of state-dependence in the black hole interior*, *Phys. Rev. D* **93** (2016) 084049 [arXiv:1503.08825] [INSPIRE].

[36] J. de Boer, R. Van Breukelen, S.F. Lokhande, K. Papadodimas and E. Verlinde, *On the interior geometry of a typical black hole microstate*, *JHEP* **05** (2019) 010 [arXiv:1804.10580] [INSPIRE].

[37] I. Kourkoulou and J. Maldacena, *Pure states in the SYK model and nearly-AdS$_2$ gravity*, arXiv:1707.02325 [INSPIRE].

[38] P. Hayden and G. Penington, *Learning the alpha-bits of black holes*, *JHEP* **12** (2019) 007 [arXiv:1807.06041] [INSPIRE].

[39] C. Bény, A. Kempf and D.W. Kribs, *Generalization of quantum error correction via the Heisenberg picture*, *Phys. Rev. Lett.* **98** (2007) 100502 [quant-ph/0608071].

[40] C. Bény, *Conditions for the approximate correction of algebras*, in *Theory of quantum computation, communication and cryptography*, Springer, Berlin, Heidelberg, Germany (2009), pg. 66.

[41] C. Bény and O. Oreshkov, *General conditions for approximate quantum error correction and near-optimal recovery channels*, *Phys. Rev. Lett.* **104** (2010) 120501 [arXiv:0907.5391].

[42] G.T. Horowitz and J.M. Maldacena, *The black hole final state*, *JHEP* **02** (2004) 008 [hep-th/0310281] [INSPIRE].

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

[44] S.W. Hawking and D.N. Page, *Thermodynamics of black holes in anti-de Sitter space*, *Commun. Math. Phys.* **87** (1983) 577 [INSPIRE].

[45] J.V. Rocha, *Evaporation of large black holes in AdS: coupling to the evaporon*, *JHEP* **08** (2008) 075 [arXiv:0804.0055] [INSPIRE].

[46] M. Van Raamsdonk, *Evaporating firewalls*, *JHEP* **11** (2014) 038 [`arXiv:1307.1796`] [ɪɴSPIRE].

[47] J. Preskill, *Quantum information and computation*, Lecture notes for physics 229, California Institute of Technology, Pasadena, CA, U.S.A. (1998).

[48] C. Akers, N. Engelhardt, G. Penington and M. Usatyuk, *Quantum maximin surfaces*, `arXiv:1912.02799` [ɪɴSPIRE].

[49] R. Bousso, Z. Fisher, S. Leichenauer and A.C. Wall, *Quantum focusing conjecture*, *Phys. Rev. D* **93** (2016) 064044 [`arXiv:1506.02669`] [ɪɴSPIRE].

[50] G. Penington, S.H. Shenker, D. Stanford and Z. Yang, *Replica wormholes and the black hole interior*, `arXiv:1911.11977` [ɪɴSPIRE].

[51] 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`] [ɪɴSPIRE].

[52] S. Abdolrahimi, D.N. Page and C. Tzounis, *Ingoing Eddington-Finkelstein metric of an evaporating black hole*, *Phys. Rev. D* **100** (2019) 124038 [`arXiv:1607.05280`] [ɪɴSPIRE].

[53] A. Strominger, *Les Houches lectures on black holes*, in *NATO advanced study institute: Les Houches summer school, session* 62*: fluctuating geometries in statistical mechanics and field theory*, (1994) [`hep-th/9501071`] [ɪɴSPIRE].

[54] D. Harlow, *Jerusalem lectures on black holes and quantum information*, *Rev. Mod. Phys.* **88** (2016) 015002 [`arXiv:1409.1231`] [ɪɴSPIRE].

[55] P.T. Landsberg and A.D. Vos, *The Stefan-Boltzmann constant in N-dimensional space*, *J. Phys. A* **22** (1989) 1073.

[56] A.R. Brown, H. Gharibyan, G. Penington and L. Susskind, *The python's lunch: geometric obstructions to decoding Hawking radiation*, `arXiv:1912.00228` [ɪɴSPIRE].

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

[58] P. Calabrese and J. Cardy, *Entanglement entropy and conformal field theory*, *J. Phys. A* **42** (2009) 504005 [`arXiv:0905.4013`] [ɪɴSPIRE].

[59] G. Penington, *Entanglement wedge reconstruction and the information paradox*, https://www.youtube.com/watch?v=1IXqdR5pAdE.

[60] D.N. Page, *Average entropy of a subsystem*, *Phys. Rev. Lett.* **71** (1993) 1291 [`gr-qc/9305007`] [ɪɴSPIRE].

[61] D.N. Page, *Particle emission rates from a black hole: massless particles from an uncharged, nonrotating hole*, *Phys. Rev. D* **13** (1976) 198 [ɪɴSPIRE].

[62] P. Hayden and G. Penington, *Approximate quantum error correction revisited: introducing the alpha-bit*, *Commun. Math. Phys.* **374** (2020) 369 [`arXiv:1706.09434`] [ɪɴSPIRE].

[63] D. Kretschmann and R.F. Werner, *Tema con variazioni: quantum channel capacity*, *New J. Phys.* **6** (2004) 26 [`quant-ph/0311037`].

[64] B. Yoshida, *Firewalls vs. scrambling*, *JHEP* **10** (2019) 132 [`arXiv:1902.09763`] [ɪɴSPIRE].

[65] A. Almheiri, D. Marolf, J. Polchinski, D. Stanford and J. Sully, *An apologia for firewalls*, *JHEP* **09** (2013) 018 [`arXiv:1304.6483`] [ɪɴSPIRE].

[66] M. Tomamichel, R. Colbeck and R. Renner, *A fully quantum asymptotic equipartition property*, *IEEE Trans. Inform. Theory* **55** (2009) 5840.

[67] P. Hayden and A. Winter, *Weak decoupling duality and quantum identification*, *IEEE Trans. Inform. Theory* **58** (2012) 4914.

[68] J.M. Maldacena, *Eternal black holes in anti-de Sitter*, *JHEP* **04** (2003) 021 [hep-th/0106112] [INSPIRE].

[69] S. Cooper, M. Rozali, B. Swingle, M. Van Raamsdonk, C. Waddell and D. Wakeham, *Black hole microstate cosmology*, *JHEP* **07** (2019) 065 [arXiv:1810.10601] [INSPIRE].

[70] 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].

[71] P. Hayden, M. Horodecki, A. Winter and J. Yard, *A decoupling approach to the quantum capacity*, *Open Syst. Inf. Dyn.* **15** (2008) 7.

[72] F. Dupuis, *The decoupling approach to quantum information theory*, arXiv:1004.1641.

[73] P. Saad, S.H. Shenker and D. Stanford, *JT gravity as a matrix integral*, arXiv:1903.11115 [INSPIRE].

[74] S. Lloyd and J. Preskill, *Unitarity of black hole evaporation in final-state projection models*, *JHEP* **08** (2014) 126 [arXiv:1308.4209] [INSPIRE].

[75] R. Bousso and D. Stanford, *Measurements without probabilities in the final state proposal*, *Phys. Rev. D* **89** (2014) 044038 [arXiv:1310.7457] [INSPIRE].

[76] D. Marolf, *The black hole information problem: past, present, and future*, *Rept. Prog. Phys.* **80** (2017) 092001 [arXiv:1703.02143] [INSPIRE].

[77] N. Bao, G. Penington, J. Sorce and A.C. Wall, *Beyond toy models: distilling tensor networks in full AdS/CFT*, *JHEP* **11** (2019) 069 [arXiv:1812.01171] [INSPIRE].

[78] N. Bao, G. Penington, J. Sorce and A.C. Wall, *Holographic tensor networks in full AdS/CFT*, arXiv:1902.10157 [INSPIRE].

[79] T. Faulkner and A. Lewkowycz, *Bulk locality from modular flow*, *JHEP* **07** (2017) 151 [arXiv:1704.05464] [INSPIRE].

[80] C.-F. Chen, G. Penington and G. Salton, *Entanglement wedge reconstruction using the Petz map*, *JHEP* **01** (2020) 168 [arXiv:1902.02844] [INSPIRE].

[81] A. Hamilton, D.N. Kabat, G. Lifschytz and D.A. Lowe, *Local bulk operators in AdS/CFT: a boundary view of horizons and locality*, *Phys. Rev. D* **73** (2006) 086003 [hep-th/0506118] [INSPIRE].

[82] A. Cavaglià, S. Negro, I.M. Szécsényi and R. Tateo, *$T\bar{T}$-deformed 2D quantum field theories*, *JHEP* **10** (2016) 112 [arXiv:1608.05534] [INSPIRE].

[83] L. McGough, M. Mezei and H. Verlinde, *Moving the CFT into the bulk with $T\bar{T}$*, *JHEP* **04** (2018) 010 [arXiv:1611.03470] [INSPIRE].

[84] P. Kraus, J. Liu and D. Marolf, *Cutoff AdS_3 versus the $T\bar{T}$ deformation*, *JHEP* **07** (2018) 027 [arXiv:1801.02714] [INSPIRE].

[85] W. Donnelly and V. Shyam, *Entanglement entropy and $T\bar{T}$ deformation*, *Phys. Rev. Lett.* **121** (2018) 131602 [arXiv:1806.07444] [INSPIRE].

[86] V. Gorbenko, E. Silverstein and G. Torroba, *dS/dS and $T\bar{T}$*, *JHEP* **03** (2019) 085 [arXiv:1811.07965] [INSPIRE].

[87] A. Almheiri, R. Mahajan and J. Maldacena, *Islands outside the horizon*, arXiv:1910.11077 [INSPIRE].

[88] 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].

[89] A. Kitaev, *A simple model of quantum holography (part 1)*, in *KITP strings seminar*, University of California, Santa Barbara, CA, U.S.A., 7 April 2015.

[90] A. Kitaev, *A simple model of quantum holography (part 2)*, in *KITP strings seminar*, University of California, Santa Barbara, CA, U.S.A., 27 May 2015.

[91] J. Maldacena and D. Stanford, *Remarks on the Sachdev-Ye-Kitaev model*, *Phys. Rev. D* **94** (2016) 106002 [arXiv:1604.07818] [INSPIRE].

[92] J. Polchinski and V. Rosenhaus, *The spectrum in the Sachdev-Ye-Kitaev model*, *JHEP* **04** (2016) 001 [arXiv:1601.06768] [INSPIRE].

[93] J.S. Cotler et al., *Black holes and random matrices*, *JHEP* **05** (2017) 118 [*Erratum ibid.* **09** (2018) 002] [arXiv:1611.04650] [INSPIRE].

[94] P. Saad, S.H. Shenker and D. Stanford, *A semiclassical ramp in SYK and in gravity*, arXiv:1806.06840 [INSPIRE].

[95] A. Almheiri and J. Polchinski, *Models of AdS$_2$ backreaction and holography*, *JHEP* **11** (2015) 014 [arXiv:1402.6334] [INSPIRE].

[96] J. Engelsöy, T.G. Mertens and H. Verlinde, *An investigation of AdS$_2$ backreaction and holography*, *JHEP* **07** (2016) 139 [arXiv:1606.03438] [INSPIRE].

[97] K. Jensen, *Chaos in AdS$_2$ holography*, *Phys. Rev. Lett.* **117** (2016) 111601 [arXiv:1605.06098] [INSPIRE].

[98] J. Maldacena, D. Stanford and Z. Yang, *Conformal symmetry and its breaking in two dimensional nearly anti-de-Sitter space*, *PTEP* **2016** (2016) 12C104 [arXiv:1606.01857] [INSPIRE].

[99] M. Ohya and D. Petz, *Quantum entropy and its use*, Springer, Berlin, Heidelberg, Germany (1993).