The entropy of bulk quantum fields and the entanglement wedge of an evaporating black hole

Ahmed Almheiri,[a] **Netta Engelhardt,**[b,c] **Donald Marolf,**[d] **Henry Maxfield**[d]

[a] *Institute for Advanced Study, Princeton, NJ 08540, USA*

[b] *Department of Physics, Princeton University, Princeton, NJ 08544, USA*

[c] *Gravity Initiative, Princeton University, Princeton NJ 08544, USA*

[d] *Physics Department, University of California, Santa Barbara, CA 93016*

E-mail: almheiri@ias.edu, nengelhardt@princeton.edu, marolf@ucsb.edu, hmaxfield@physics.ucsb.edu

Abstract: Bulk quantum fields are often said to contribute to the generalized entropy $\frac{A}{4G_N} + S_{\text{bulk}}$ only at $O(1)$. Nonetheless, in the context of evaporating black holes, $O(1/G_N)$ gradients in S_{bulk} can arise due to large boosts, introducing a quantum extremal surface far from any classical extremal surface. We examine the effect of such bulk quantum effects on quantum extremal surfaces (QESs) and the resulting entanglement wedge in a simple two-boundary $2d$ bulk system defined by Jackiw-Teitelboim gravity coupled to a 1+1 CFT. Turning on a coupling between one boundary and a further external auxiliary system which functions as a heat sink allows a two-sided otherwise-eternal black hole to evaporate on one side. We find the generalized entropy of the QES to behave as expected from general considerations of unitarity, and in particular that ingoing information disappears from the entanglement wedge after a scambling time $\frac{\beta}{2\pi} \log \Delta S + O(1)$ in accord with expectations for holographic implementations of the Hayden-Preskill protocol. We also find an interesting QES phase transition at what one might call the Page time for our process.

Contents

1 Introduction

A key concept in our current understanding of holographic dualities is the entanglement wedge W_A of a bulk spacetime associated with a given region A in the dual holographic field theory [1–3]. At the level of classical bulk physics, W_A is obtained by first constructing the associated Hubeny-Rangamani-Takayanagi (HRT) surface [4], which is the minimal-area codimension-2 bulk extremal surface homologous to A in an

appropriate sense [5]. The physics in the wedge W_A can then be reconstructed from field theory degrees of freedom in A [6–9].

Key steps in the above arguments rely on identifying the area $\mathcal{A}$ of the HRT surface as $4G_N$ times the von Neumann entropy of the field theory degrees of freedom in A [4, 10–13]. Here G_N is the bulk Newton constant and the bulk is treated classically. But at the quantum level the von Neumann entropy of the field theory degrees of freedom in A is instead the so-called generalized entropy of a bulk surface X [14], which for Einstein-Hilbert gravity coupled to $O(1)$ bulk quantum fields may be written $S_{\rm gen} = \frac{\mathcal{A}(X)}{4G_N} + S_{\rm bulk}(X)$, where $S_{\rm bulk}(X)$ is a von Neumann entropy of bulk quantum fields on one side of X. As explained in [15] and partially verified in [13], this means that bulk quantum effects should move the boundary of W_A to a so-called quantum extremal surface X extremizing $S_{\rm gen}$ as defined by $\mathcal{A}[X]$ and by the von Neumann entropy of bulk quantum fields between X and A. As before, X should satisfy the homology constraint and, when there is more than one such quantum extremal surface, we should choose the one minimizing $S_{\rm gen}$.

One often thinks of such quantum corrections as being small. In many contexts $S_{\rm bulk}$ is indeed $O(1)$ and the relevant quantum and classical extremal surfaces nearly coincide. In such contexts one should also consider corrections to the classical formula $\frac{\mathcal{A}}{4G_N}$ associated with higher derivative corrections to the gravitational effective action [16–20]. But in contexts involving long times and/or long distances, secular effects can cause $S_{\rm bulk}$ to grow and in some cases to become of order $1/G_N$, and large boosts can cause sharp gradients even when $S_{\rm bulk}$ remains $O(1)$. The last of these, which we investigate here, is particularly natural in the context of black hole evaporation, where the semi-classical Hawking effect leads to bulk entropy comparable to the Bekenstein-Hawking entropy of the original black hole [21, 22] and large boosts arise naturally from time evolution.

We study a simple model where such $O(1/G_N)$ gradients of bulk entropy can be calculated in detail and the ensuing effects on quantum extremal surfaces (QESs) and entanglement wedges can be studied. We consider a standard two-sided AdS_2 black hole in Jackiw-Teitelboim (JT) gravity coupled to a 1+1 CFT in the Hartle-Hawking state. With reflecting boundary conditions at AdS infinity, the state is independent of time. But at a finite time we couple the right boundary of our system to an auxiliary system B, which functions as a bath, or heat sink. We take B to (1) be a copy of the same 1+1 CFT on the right half of Minkowski space and (2) begin in its own vacuum. The coupling is such that, after a short transition, the right boundary is fully transparent. In effect, the coupling simply glues the origin of the auxiliary Minkowski space to the right AdS_2 boundary; see figure 1. This leads to evaporation on the right of the two-sided AdS_2 black hole.

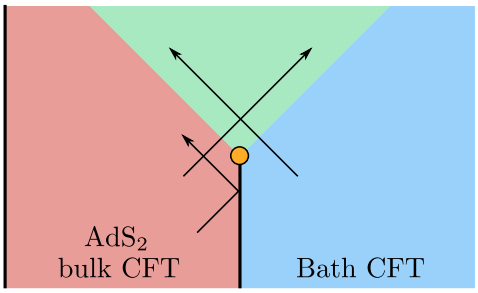

Figure 1. Our two-sided AdS_2 system initially has reflecting boundary conditions (solid vertical lines) on its right boundary. An independent copy B of our CFT on the right half of Minkowski space which will play the role of the bath also begins with reflecting boundary conditions. At some finite time (orange dot), the right-AdS_2 boundary conditions become transparent, coupling the AdS_2 CFT to the Bath CFT.

Because the AdS_2 system is no longer isolated, bulk von Neumann entropies depend on a choice of Cauchy slice, or at least a choice of where such slices meet the right AdS_2 boundary. In this sense, the QES of the right boundary becomes time-dependent. The time-dependent QESs may be viewed as a proxy for what one would find if one turned off the coupling at the given time, used the data on the stated Cauchy surface as initial data for a new AdS_2 bulk, and computed the QES in the resulting isolated spacetime. The isolated QES and the proxy QES in the coupled spacetime will coincide up to corrections associated with the details of how the coupling is switched off. This setting and aspects of JT gravity are reviewed in section 2, while section 3 presents initial studies of the matter sector.

The proxy QESs are studied in section 4. Although we consider only standard perturbative semiclassical bulk physics, tracking the proxy QES and computing $S_{\rm gen}$ as a function of boundary time reproduces features one would expect from general considerations of fully unitary evolution. In particular, the Page time, when the fine-grained von Neumann entropy of the black hole saturates at the coarse-grained thermodynamic entropy, is marked by a phase transition where the quantum extremal surface jumps. Thereafter, the location of the quantum extremal surface gives a quantum geometric realization of the Hayden-Preskill protocol [23], as described holographically in [24]. For the convenience of the reader, the technical results are then summarized in section 4.6.

An important part of the above description is the gap between the QESs X_L and X_R associated with the left and right boundaries of AdS_2, and this gap is discussed further in section 5. We close with further discussion in section 6, which in particular

describes analogous effects in cases where black holes evaporate more completely. The final interpretation of such results is unclear, but will clearly fuel further discussion of black hole information puzzles, firewalls, state-dependence, and related issues; see e.g. [25, 26] for recent reviews and [24] for further recent discussion.

Note to reader: While this work was underway we learned that related results were independently found in the work [27] which will appear on the arxiv simultaneously.

1.1 Holographic Hayden-Preskill

Before proceeding with the main paper in section 2, we pause to give a brief review of the Hayden-Preskill protocol [23], and expectations for its holographic realization [24]. The protocol considers an old black hole past the Page time which is maximally entangled with its early radiation. Assuming the black hole is governed by a sufficiently scrambling internal Hamiltonian, abstract quantum information reasoning is used to show that information thrown into this black hole would be recoverable from the radiation in a relatively short time compared to the black hole lifetime.

The protocol can be described as follows: Consider throwing some information m in the state $|i\rangle_m$ into an old black hole $\mathcal{B}$ maximally entangled with the early radiation $\mathcal{E}$ in the state $|\psi\rangle_{\mathcal{B}\mathcal{E}}$. After allowing for the black hole interior to scramble via its own internal unitary dynamics governed by some unitary $U_{m\mathcal{B}}$, the black hole is allowed to evaporate into some new radiation $\mathcal{L}$ with the remaining black hole given by $\mathcal{B}'$. This process is described by

$$|i\rangle_m|\psi\rangle_{\mathcal{B}\mathcal{E}} \to U_{m\mathcal{B}}|i\rangle_m|\psi\rangle_{\mathcal{B}\mathcal{E}} \equiv |\Psi_i\rangle_{m\mathcal{B}\mathcal{E}} = |\Psi_i\rangle_{\mathcal{B}'\mathcal{L}\mathcal{E}}. \tag{1.1}$$

The last equality just comes from the identification of $m\mathcal{B}$ and $\mathcal{B}'\mathcal{L}$. The result of [23] is that it is sufficient that $U_{m\mathcal{B}}$ to be drawn from a unitary 2-design[1] for the message to be recoverable from the radiation subsystem $\mathcal{L}\mathcal{E}$, in the sense that

$$\forall i, \quad \exists V_{\mathcal{L}\mathcal{E}} \quad \text{s.t.} \quad V_{\mathcal{L}\mathcal{E}}|\Psi_i\rangle_{\mathcal{B}'\mathcal{L}\mathcal{E}} = |i\rangle_{l_1} \otimes |\chi\rangle_{\mathcal{B}'l_2}, \tag{1.2}$$

for some fixed $|\chi\rangle$, and where $l_1, 2$ are some Hilbert space factors of $\mathcal{L}\mathcal{E}$.

This protocol predicts an interesting time scale for when the message appears in the Hawking radiation, which stems from the assumption that the black hole internal unitary dynamics is given by a unitary 2-design. These circuits have logarithmic depth in the number of qubits of the black hole and thus naturally suggest the time scale

$$t_{HP} \sim \frac{1}{T} \log S_{BH}, \tag{1.3}$$

[1] It is defined as that which coincides with the Harr measure up to second moments in $U_{ij}U^\dagger_{kl}$.

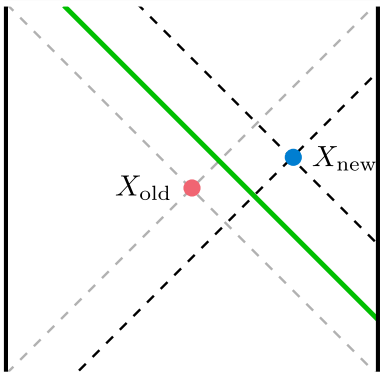

Figure 2. After the evaporation of the right CFT R into the left L, the quantum extremal surface moves in a spacelike direction towards the right boundary from X_{old} to X_{new}. A message sent into R in the past will escape the new entanglement wedge of R and enter that of L.

where T and S_{BH} are the temperature and entropy of the black hole respectively. This time, called the scrambling time, places a lower bound on the time needed before the message appears in the radiation. We will see in section 4 how this timescale naturally arises in a precise form from the evolution of the QES of an evaporating black hole.

As stated, the Hayden-Preskill protocol is a general statement about scrambling systems and would therefore naturally apply to the case of two entangled holographic CFTs. The setup of this protocol in the holographic context was recently discussed in [24], where the analogue of flat space post-Page time black hole is two entangled holographic CFTs, $L \otimes R$, in the thermofield double dual to the eternal black hole in AdS.

One obvious objection at this point is that the eternal black hole does not evaporate, so any information injected on one side will remain there eternally. We can mimic black hole evaporation by extracting energy from R and dumping it into L; the right and left play the role of an old black hole and its early radiation respectively. Our goal is to study the evolution of the quantum extremal surface in the evaporating spacetime as a proxy for understanding, at each time, properties that the bulk-boundary dictionary would have if the coupling were turned off. Applying the Hayden-Preskill reasoning in this case would say that information sent into R should become reconstructable in L after a few scrambling times.

The general evaporation protocol will be as follows:

1. Start with two entangled CFTs in the thermofield double state with temperature above the Hawking-Page transition dual to the eternal black hole.

2. Introduce an auxiliary bath system B, taken to be a large system in the vacuum, which we couple to R and allow the Hawking radiation to be extracted from R to B. This is provided in AdS by imposing absorbing boundary conditions on the asymptotic boundary of R.

3. This Hawking radiation is then transferred into L. One can imagine transferring this information into any of the many low occupation modes on the left. This will excite the state of the quantum fields on the left exterior.

The end result will be a new pure state of $L \otimes R$ where the ADM energy on the right is lower than that on the left (and lower than the initial energy on the right), with smaller entanglement between the two.

We will explain in detail in this paper how this protocol achieves Hayden-Preskill by inducing a motion of the QES surface away from the bifurcate horizon, causing the information injected into R at early times to escape the entanglement wedge of that side. The rough idea is shown in figure 2. Since we expect that the modified state of LR still exhibits complementary recovery, the inserted message enters the entanglement wedge of L, and can therefore be decoded from it, in the sense of (1.2).

The scrambling time for near extremal black holes found in [28] is controlled by the ADM energy E above the ground state and the energy of the perturbation δE thrown into the black hole, assumed to be much smaller than E, as

$$t_{scr} = \alpha_S \frac{\beta}{2\pi} \log \frac{E}{\delta E}, \tag{1.4}$$

for some constant α_S. This result indicates the Hayden-Preskill time should be

$$t_{HP} = \alpha_{HP} \frac{\beta}{2\pi} \log(S - S_0), \tag{1.5}$$

for a small message with $\delta E \sim E/(S - S_0)$, and for some constant α_{HP} which a priori could be distinct from α_S. The realization of the Hayden-Preskill protocol in our context will be exhibited by the lag of the QES of the state at time t by an amount t_{HP} in null ingoing time; this would ensure that messages thrown in prior to $t - t_{HP}$ would escape the entanglement wedge. We confirm this expectation and determine the values of α_{HP} and α_S for systems dual to JT gravity.

We wish to emphasize the difference between this described protocol and the recent story of making a wormhole traversable via a double trace deformation [29, 30]. Traversability is achieved by violating the Averaged Null Energy Condition (ANEC) on the horizon, which provides a message falling into the horizon sufficent time advance that it emerges into the other asymptotic region of the eternal black hole. This is in

contrast with the proposal of this paper where it is the evolution of the dictionary under the evaporation protocol that renders the message recoverable from the other boundary CFT.

Second, the traversable wormhole protocol takes advantage of the careful local correlations in the TFD between the two CFTs at $t = 0$ (or by boost invariance - opposite times) and picks a deformation with large connected expectation value. This sensitivity to the state implies a sensitivity to the time at which message is thrown in. In particular, it works best for messages thrown in at around the scrambling time prior to turning on the interaction. A message sent in too early would spoil the delicate correlations in the TFD, thereby ruining the efficacy of the deformation. In the bulk, this is interpreted as the failure of the eikonal approximation of scattering between the message and the negative stress tensor, which precludes the necessary time advance for traverability [30]. Sent in late, the message simply doesn't get enough of a time advance to make it through. The evaporation protocol in this paper does not suffer from this issue, and as we will see, all messages thrown into the black hole will eventually appear in the entanglement wedge of the complement after a scrambling time.

2 Evaporating Near-Extremal Black Holes in JT Gravity

2.1 Review of JT Gravity

We will study the evolution of the minimal quantum extremal surface (QES) in an evaporating black hole in JT gravity coupled to conformal matter. The dynamics of this theory are governed by the Lorentzian action $I = I_0 + I_G + I_M$ with

$$I_0 = \frac{\phi_0}{16\pi G_N}\left[\int_{\mathcal{M}} d^2x\sqrt{-g}\ R + 2\int_{\partial M} K\right], \tag{2.1}$$

$$I_G = \frac{1}{16\pi G_N}\left[\int_{\mathcal{M}} d^2x\sqrt{-g}\ \phi(R+2) + 2\int_{\partial M}\ \phi_b K\right], \tag{2.2}$$

$$I_M = I_{CFT}[g]. \tag{2.3}$$

The dynamics of this model are especially simple. The gravitational action can be thought of as the dimensional reduction of a higher dimensional theory describing the s-wave sector of the near horizon limit of near extremal black holes [31–34]. From this perspective, the area of the transverse space in the higher dimensional theory becomes the dilaton $\phi_0 + \phi$, thereby implicitly imposing the restriction $\phi_0 \gg \phi$. The action I_0 is a purely topological term and provides the extremal entropy of the black hole $\phi_0/4G_N$.

The remaining gravitational dynamics are governed by the action I_G, which is not topological because ϕ is dynamical. This action is easily solved by integrating out the

dilaton along an imaginary contour which imposes the constraint on the spacetime to have constant negative curvature via the delta function

$$\delta(R+2), \tag{2.4}$$

which requires that the two dimensional metric is locally AdS_2. In Poincaré coordinates this is

$$ds^2 = \frac{-dt^2+dz^2}{z^2} = -\frac{4dx^+dx^-}{(x^+-x^-)^2}, \quad x^\pm = t \pm z. \tag{2.5}$$

As we review below, this AdS_2 space should be thought of as an 'ambient' rigid space of which the actual physical spacetime is a patch [31, 32, 35, 36]. Varying the action with respect to the metric g yields the constraints and equation of motion that couple the bulk CFT to the dilaton:

$$2\partial_{x^+}\partial_{x^-}\phi + \frac{4}{(x^+-x^-)^2}\phi = 16\pi G_N T_{x^+x^-}, \tag{2.6a}$$

$$-\frac{1}{(x^+-x^-)^2}\partial_{x^+}\left((x^+-x^-)^2\partial_{x^+}\phi\right) = 8\pi G_N T_{x^+x^+}, \tag{2.6b}$$

$$-\frac{1}{(x^+-x^-)^2}\partial_{x^-}\left((x^+-x^-)^2\partial_{x^-}\phi\right) = 8\pi G_N T_{x^-x^-}. \tag{2.6c}$$

We work in the limit where the gravitational sector can be treated semiclassically, so we may replace the stress tensors with their expectation values, $T_{ab} = \langle T_{ab}\rangle$.

It is often convenient to express JT gravity as the dynamics of the so-called 'boundary particle' [31, 32, 36]. This is simply a reparametrization between the bulk Poincaré time near the boundary t and the physical boundary time u. The location of this physical boundary is specified by the boundary condition on the bulk fields

$$g_{uu}\big|_{bdy} = \frac{1}{\epsilon^2} = \frac{-t'^2+z'^2}{z^2}, \ \phi = \phi_b = \frac{\bar{\phi}_r}{\epsilon}, \tag{2.7}$$

where g_{uu} is the time-time component of the metric near the boundary along the physical boundary time u. The last equality in (2.7) indicates that we are interested in large ϕ_b ($\phi_b \sim 1/\epsilon$) with fixed constant coefficient $\bar{\phi}_r$. With this choice, the JT action reduces to a boundary term given by

$$S_G = \frac{1}{8\pi G_N}\int_{\partial M}\phi_b K \to \frac{\bar{\phi}_r}{8\pi G_N}\int du\{f(u),u\}, \tag{2.8}$$

where $t = f(u)$ is a diffeomorphism giving Poincaré time t in terms of boundary proper time u. This is the Schwarzian action, which is invariant under $SL_2(\mathbb{R})$ transformations

of the trajectory of the 'boundary particle' $t = f(u)$, as required by the isometries of the rigid AdS_2 spacetime. From this description it is easy to compute the ADM energy of the spacetime [31, 32], defined as the Noether charge under physical time translations $u \to u + \delta u$

$$E(u) = -\frac{\bar{\phi}_r}{8\pi G_N}\{f(u), u\}. \tag{2.9}$$

Using this diffeormophism, we can construct natural coordinates $y^\pm$ defined by $x^\pm = f(y^\pm)$, in which the metric becomes more complicated,

$$ds^2 = -\frac{4f'(y^+)f'(y^-)dy^+dy^-}{(f(y^+) - f(y^-))^2}, \tag{2.10}$$

but the cutoff is simpler, at constant $\frac{y^+ - y^-}{2} = \epsilon$.

The vacuum solutions, with vanishing stress-tensor expectation value[2], have dilaton profile

$$\phi = 2\bar{\phi}_r \frac{1 - (\pi T_0)^2 x^+ x^-}{x^+ - x^-} \tag{2.11}$$

up to gauge transformations, which represents an eternal black hole with two asymptotic boundaries and temperature T_0 [35]. The associated reparameterization is

$$f(u) = \frac{1}{\pi T_0}\tanh(\pi T_0 u), \tag{2.12}$$

and using this to transform to y coordinates (which cover the exterior of the black hole) the metric and dilaton take the manifestly static form

$$ds^2 = \frac{-4dy^+dy^-}{\frac{1}{(\pi T_0)^2}\sinh^2\left[\pi T_0(y^+ - y^-)\right]}, \quad \phi = 2\bar{\phi}_r \pi T_0 \coth\left[\pi T_0(y^+ - y^-)\right]. \tag{2.13}$$

The boundary particle meets the AdS_2 boundary at two locations:

$$\text{for} \quad u \to \infty, \quad x^+ = t = \frac{1}{\pi T_0}, \tag{2.14a}$$

$$\text{for} \quad u \to -\infty, \quad x^- = t = -\frac{1}{\pi T_0}. \tag{2.14b}$$

[2]For the case of conformal matter we are considering, the trace of the stress tensor is a constant determined by the conformal anomaly $T^\mu_\mu = \frac{c}{24\pi}R$, which can be absorbed into the extremal value of the dilaton: $\phi_0 \to \phi_0^{Ren} = \phi_0 + \frac{cG_N}{3}$.

Plugging the reparameterization (2.12) into the formula for the energy we find

$$E(u) = -\frac{\bar{\phi}_r}{8\pi G_N}\left\{\frac{1}{\pi T_0}\tanh\left(\pi T_0 u\right), u\right\} = \frac{\pi\bar{\phi}_r}{4G_N}T_0^2 \equiv E_0 \tag{2.15}$$

which is the expected leading scaling with temperature of the energy of a near extremal black hole[3].

2.2 Evaporation

Starting with the static solution above, we will couple the right boundary to a large external heat bath B at zero temperature, thereby extracting the Hawking radiation and evaporating the right side of wormhole. We describe an explicit model for this coupled evolution in section 3, here discussing the consequences which are pertinent to the bulk dynamics, namely the resulting energy-momentum transfer into the black hole.

First, an important transient effect occurs when coupling the right boundary to this external system, namely an initial injection of positive energy into the black hole. This is required to satisfy the ANEC along its horizon, and to prevent the wormhole from becoming traversable. While the presence of this positive energy is required by consistency, its precise value depends on the details of the system and bath and the precise coupling between the two. We will denote this initial positive energy increase as E_S, and find a lower bound on its value.

After this initial 'shock' of energy, the energy of the black hole begins to be transferred into the bath via the Hawking radiation. In section 3 we give a very explicit model for coupling, in which we compute the resulting stress-tensor expectation value. This relies on choosing conformally invariant matter and boundary conditions, so that the stress tensor expectation value is determined by the conformal anomaly. For the energy-momentum expectation value, it gives the same result as the analysis of [32], which used a model of perfect absorption of outgoing Hawking quanta at the boundary. The result (derived in (3.23)) is that after the shock, the ingoing stress-tensor expectation value vanishes in the flat metric $-dy^+dy^-$, which via the conformal anomaly gives a flux of negative energy in the physical metric (2.10):

$$\langle T_{x^-x^-}(x^-)\rangle = E_S\,\delta(x^-) - \frac{c}{24\pi}\{y^-, x^-\}\Theta(x^-) \tag{2.16}$$

This is valid after scaling the Poincaré coordinates to set $f'(0) = 1$, and we recall that $x^- = f(y^-)$. We can alternatively rewrite the last term using the inversion identity for the Schwarzan, $\{y^-, x^-\} = -f'(x^-)^{-2}\{f(x^-), x^-\}$.

[3]This follows if we assume that the entropy of a near extremal black hole is analytic near $T = 0$. Using the first law of thermodynamics implies that energy must be quadratic in T.

As discussed in [32], this result can be used to solve for the energy of the black hole as a function of boundary time. Varying $I_G + I_{CFT}[g]$ with respect to boundary time yields the energy balance equation

$$\partial_u E(u) = f'(u)^2 \left(T_{x^-x^-} - T_{x^+x^+}\right), \tag{2.17}$$

equating the change in energy to the ingoing flux minus outgoing flux. Using this and the expression for energy in terms of the Schwarzian (2.9), for positive times we find the differential equation

$$-\frac{\bar{\phi}_r}{8\pi G_N}\partial_u\{f(u),u\} = \frac{c}{24\pi}\{f(u),u\}, \implies \{t(u),u\} \propto e^{-ku} \tag{2.18}$$

where

$$k = \frac{c}{12}\frac{4G_N}{\bar{\phi}_r} \ll 1. \tag{2.19}$$

Putting in the initial energy and the perturbation due to turning on the interaction, the energy as a function of time is found to be

$$E(u) = \Theta(-u)E_0 + \Theta(u)E_1 e^{-ku}, \tag{2.20}$$

where $E_1 \equiv E_0 + E_S$, and E_S is the positive energy due to turning on the coupling between the system and bath. For small positive time, we have a black hole with new temperature T_1 satisfying

$$E_1 \equiv \frac{\pi\bar{\phi}_r}{4G_N}T_1^2. \tag{2.21}$$

We can put a bound on the magnitude of E_S by requiring that the new event horizon lies outside the original horizon, so the wormhole does not become traversable. Since there is no interaction between the left and right boundaries, a traversable wormhole would violate boundary causality. Moreover, in JT gravity traversable wormholes would require violations of the ANEC [37, 38] which, in our context, are forbidden by extending the results of [39] to Killing horizons in curved space as described in that reference. The new horizon is at $x^+ = t_\infty$, where $t_\infty = \lim_{u\to\infty} f(u)$ is the Poincaré time at which the boundary particle reaches the boundary, so we require $t_\infty < \frac{1}{\pi T_0}$.

For $u < 0$, the reparameterization f is given by the black hole solution (2.12). For $u > 0$, we must solve the differential equation

$$\{f(u),u\} = 2(\pi T_1)^2 e^{-ku}, \qquad f(0) = 0,\ f'(0) = 1,\ f''(0) = 0, \tag{2.22}$$

where the initial conditions come from matching to the $u < 0$ solution at $u = 0$. Explicitly, the solution is

$$f(u) = \frac{1}{\pi T_1} \frac{-K_0\left[\frac{2\pi T_1}{k}\right] I_0\left[\frac{2\pi T_1}{k} e^{-ku/2}\right] + I_0\left[\frac{2\pi T_1}{k}\right] K_0\left[\frac{2\pi T_1}{k} e^{-ku/2}\right]}{K_1\left[\frac{2\pi T_1}{k}\right] I_0\left[\frac{2\pi T_1}{k} e^{-ku/2}\right] + I_1\left[\frac{2\pi T_1}{k}\right] K_0\left[\frac{2\pi T_1}{k} e^{-ku/2}\right]}, \tag{2.23}$$

which gives

$$t_\infty = \frac{1}{\pi T_1} \frac{I_0\left[\frac{2\pi T_1}{k}\right]}{I_1\left[\frac{2\pi T_1}{k}\right]} = \frac{1}{\pi T_1} + \frac{k}{4(\pi T_1)^2} + O(k^2), \tag{2.24}$$

where we have expanded for $k \sim G_N \ll 1$. Our causality requirement $t_\infty < \frac{1}{\pi T_0}$ then gives a lower bound on E_S, which we solve or at leading order in k:

$$(\pi T_1)^2 = (\pi T_0)^2 + \frac{4\pi G_N}{\bar{\phi}_r} E_S \implies E_S > \frac{c}{24} T_0 + O(k) \tag{2.25}$$

An injection of positive energy on the thermal scale is required to maintain boundary causality.

The complicated Bessel function expression for the reparameterization is only really necessary at very late times, $u \sim k^{-1} \log k$, by which time the black hole is so close to extremality that the semiclassical description breaks down, since we no longer have a parametrically large non-extremal entropy. At times of order k^{-1}, when the black hole has evaporated an order one fraction of its mass but remains sufficiently far from extremality, we can approximate f by a simpler form. For this, we use the asymptotic formula

$$\frac{K_n(z)}{\pi I_n(z)} \sim e^{-2z}\left(1 + \frac{4n^2 - 1}{4z} + O(z^{-2})\right), \tag{2.26}$$

expanding at small k with fixed uk. The result is

$$\log\left(\frac{t_\infty - f(u)}{2t_\infty}\right) \sim -\frac{4\pi T_1}{k}\left(1 - e^{-\frac{k}{2}u}\right) + O(ke^{\frac{k}{2}u}), \tag{2.27}$$

where we have included the correction term relevant at late times; a different correction becomes important at early times, when u is of order one.

Differentiating, we find

$$\frac{1}{t_\infty - t} f'(u) \sim 2\pi T_1 e^{-\frac{k}{2}u} + O(k^2 e^{\frac{k}{2}u}). \tag{2.28}$$

Taking further derivatives, we can verify that this approximation solves the required equation for the Schwarzian to the specified order. Using these estimates along with the inversion formula for the Schwarzian, we can estimate the stress tensor at these times via

$$\{u, t\} \sim \frac{1}{2(t_\infty - t)^2}\left(1 + O(k^2 e^{-ku})\right), \tag{2.29}$$

where $t = f(u)$.

With a given stress tensor, the constraints and equation of motion for the metric can be solved in terms of an integral of the stress tensor [35]. To the future of the shock $x^- > 0$, this solution can be written as

$$\phi = 2\phi_r \frac{1 - (\pi T_1)^2 x^+ x^- + \frac{1}{2} k I(x^+, x^-)}{x^+ - x^-}, \tag{2.30}$$

where I is the integral

$$I(x^+, x^-) = \int_0^{x^-} dt (x^+ - t)(x^- - t)\{u, t\}. \tag{2.31}$$

Here, we have used the form (2.16) to write the ingoing stress tensor in terms of the Schwarzian, and $u = f^{-1}(t)$. To find quantum extremal surfaces, we will be interested in the variation of the dilaton. For this, we can use the integral expressions

$$(x^+ - x^-)^2 \partial_\pm \left(\frac{I}{x^+ - x^-} \right) = \mp \int_0^{x^-} dt (x^\mp - t)^2 \{u, t\}. \tag{2.32}$$

For early times, using the approximation $f(u) \sim \frac{1}{\pi T_1} \tanh(\pi T_1 u)$, we have

$$\{u, t\} \sim \frac{2(\pi T_1)^2}{(1 - (\pi T_1 t)^2)^2}, \tag{2.33}$$

which can be used, along with the integral expressions to obtain simple explicit expressions for I and derivatives of the dilaton.

For later times, we can use (2.29), which can be seen to match with (2.33) at intermediate times; at later times, it will prove vitally important that the double pole is shifted to t_∞, which is corrected from $\frac{1}{\pi T_1}$ by a power series (2.24) in k. We will follow through these calculations as required in section 4.

3 The matter sector

In our model, the matter sector is independent of the dilaton, coupling to it only through the constraints. We can therefore treat it as a quantum field theory on a fixed AdS_2 background. In this section we describe our model for the matter and its coupling to the auxiliary system collecting the Hawking radiation. We then compute the quantities relevant for our purposes, namely the stress tensor expectation value, which determines the dynamics of the dilaton, and the entropies of subsystems.

Explicit calculations are made possible by choosing a conformally invariant matter theory, with conformally invariant boundary conditions at the asymptotics of AdS_2

before turning on the coupling to the bath. For example, we could choose free massless fields with reflecting boundary conditions.

We will use Poincaré coordinates for AdS_2:

$$ds^2 = \frac{-dt^2 + dz^2}{z^2} = -\frac{4dx^+dx^-}{(x^+ - x^-)^2} = \frac{4dxd\bar{x}}{(x+\bar{x})^2}, \qquad \begin{cases} x^+ = t + z = \bar{x} \\ x^- = t - z = -x. \end{cases} \tag{3.1}$$

The coordinates $x, \bar{x}$ are useful for describing the preparation of our state by a path integral in a Euclidean spacetime with Euclidean time $\tau = it$, so $x = z + i\tau$ and $\bar{x}$ is its complex conjugate. After conformal transformation, the Hartle-Hawking state on AdS_2 is given by the vacuum on the half-line $z > 0$, where we choose conformally invariant boundary conditions at $z = 0$.

3.1 Coupling to the bath

Starting with the matter in this state at time $t = 0$, we want to allow the black hole to evaporate via a coupling to an auxiliary system which acts as a bath to collect the Hawking radiation. Here, we simply choose the bath to be another half-line supporting the same CFT as the bulk matter theory, also initially in the vacuum with the same conformally invariant boundary condition. At $t = 0$, we remove the boundary between AdS_2 and the heat bath, allowing matter to move freely between the two.

However, we must be slightly careful when we couple the systems, because we want to match the time evolution in the bath with the physical time evolution of the gravitational system. The physical time u does not match the Poincaré time t; rather they are related by a diffeomorphism $t = f(u)$, where may choose $f(0) = 0$. We must couple the bath at time u to the same physical time at the boundary of AdS_2, which corresponds to Poincaré time $t = f(u)$.

We can simplify the time evolution by using local conformal symmetry. Namely, we change to coordinates $y, \bar{y}$ such that $x = f(y)$, and $\bar{x} = f(\bar{y})$ on AdS_2, in which the metric is more complicated but the boundary coordinate time corresponds to physical time u:

$$ds^2 = \frac{4f'(y)f'(\bar{y})dyd\bar{y}}{(f(y) + f(\bar{y}))^2} = \Omega_y^{-2} dyd\bar{y} \tag{3.2}$$

For the AdS boundary to lie at $y + \bar{y} = 0$, and for $\frac{\bar{y}-y}{2}$ to correspond to the physical time u there, we define f for $y < 0$ by extending it as an odd function, $f(-y) = -f(y)$. We should regard this as preparing a time-reversal invariant state at $t = 0$, which we can evolve symmetrically in either direction with the coupled Hamiltonian; this is different from the physical time evolution, which has a decoupled Hamiltonian and different function f in the past.

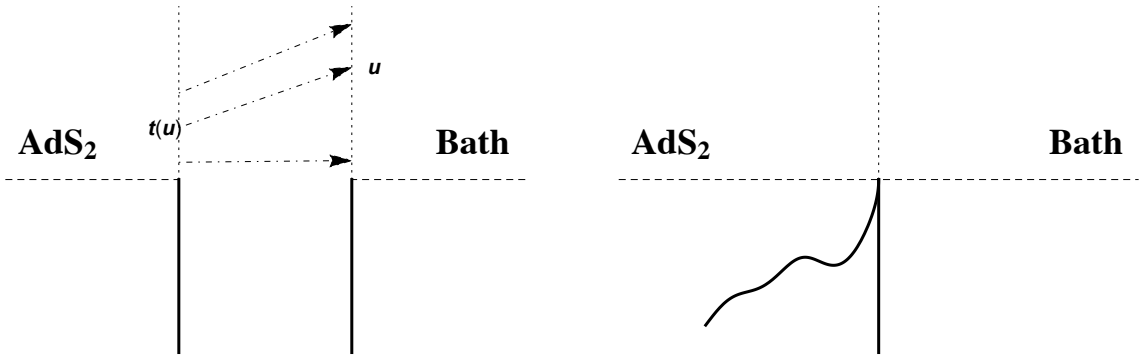

Figure 3. Left: we prepare the state at $t = 0$ (dashed line) in both AdS_2 and bath in the half-line vacuum, given by the Euclidean path integral in the lower half, but for subsequent time evolution must identify boundary times with the function $u = f(t)$. Right: after making a diffeomorphism in the AdS half, the state is prepared by a path integral with deformed boundary, but time evolution of the coupled system is given simply by the Hamiltonian of the CFT on the line. The lower and upper parts of the right diagram represent the Euclidean and Lorentzian pieces of the path integral in the y coordinates.

We can now make a Weyl transformation using Ω_y to the flat metric $dyd\bar{y}$, compute in that metric, and transform back at the end. We take the bath to live in the left half-line $y + \bar{y} < 0$, with the state prepared by Euclidean path integral on the left half-plane $\mathrm{Re}\, y < 0$. The combined evolution of this coupled system and bath is then implemented by the usual CFT Hamiltonian on the line, and identifies times in the desired way. The price we pay is that the state on the right half-line $y > 0$ is a complicated Virasoro descendant of the half-line vacuum, which we can think of as prepared by a path integral in the Euclidean section with a boundary of some complicated shape, as illustrated in figure 3.

We note here that the y coordinate will not cover the entire Poincaré patch, but only a Rindler patch $x, \bar{x} < f(\infty)$, since $f(y)$ remains finite as $y \to \infty$. The Euclidean path integral preparing the state on the half-line $y \in \mathbb{R}_+$ will have the topology of a cylinder. A simple example is for $f(u) = \frac{1}{\pi T} \tanh(\pi T u)$, which identifies u with Rindler time; the preparation of the state on AdS_2 is then by Euclidean path integral on a half-cylinder, with y periodically identified in imaginary time with period T^{-1}, giving a thermal state of matter fields as expected.

3.2 Mapping to the half-plane

We now have a description of our initial state on the coupled system and bath, which is characterised by a few simple properties. It is time-reflection symmetric, and a descen-

dant of the vacuum state on the half-line. This follows because it is prepared by the Euclidean path integral on the right of figure 3, over a simply connected space with a single boundary. Consequently, there is a diffeomorphism taking the initial Cauchy surface (including both the $t = 0$ slice of AdS_2 and the bath) to the half-line parameterised by a coordinate $w \in [0, \infty)$, which can be used to map our state to the half-line vacuum. Specifically, our state becomes the vacuum in the half Minkowski space $w + \bar{w} > 0$ with lightcone coordinates $w, \bar{w}$ (where we apply the same diffeomorphism to define left- and right-moving coordinates), after applying a Weyl transformation so that the metric becomes $dwd\bar{w}$. Calculation of correlation functions and entropies is then reduced to a calculation in the half-space vacuum, along with a Weyl transformation to the physical metric $ds^2 = \Omega_w^{-2} dwd\bar{w}$.

To identify the diffeomorphism to the new w coordinate, we can use the one-point function of the stress tensor, noting that it vanishes in the half-line vacuum, so $\langle T_{ww}(w)\rangle = 0$ (the subscripts denoting the metric we are working in as well as the coordinates, as is conventional in two-dimensional CFT). In AdS_2 at $t = 0$, the one-point function is zero in the physical metric, and also in the metric $dxd\bar{x}$ since the Weyl anomaly between AdS_2 and flat space in Poincaré coordinates vanishes; in the bath, the one-point function is zero in the metric $dyd\bar{y}$:

$$\langle T_{xx}(x)\rangle = 0 \quad (x > 0), \qquad \langle T_{yy}(y)\rangle = 0 \quad (y < 0) \tag{3.3}$$

Now, under most diffeomorphisms the stress tensor picks up an anomaly from the associated Weyl transformation:

$$\left(\frac{dw}{dx}\right)^2 \langle T_{ww}\rangle = \langle T_{xx}\rangle + \frac{c}{24\pi}\{w, x\} \tag{3.4}$$

Note that we use a standard normalization of the stress tensor, which differs from the common convention in the two dimensional CFT literature by a factor of -2π. The only diffeomorphisms for which the anomaly is absent, and hence which preserve the vanishing of the stress tensor one-point function, are the Möbius maps. We therefore find that for the part of the initial data slice in AdS_2, w is a Möbius map of $x = f(y)$, and for the part in the bath, w is a Möbius map of $y = f^{-1}(x)$. We can now choose the AdS_2 region $x > 0$ to map to $w \in (0, w_0)$ and the bath region $y < 0$ to map to (w_0, ∞), from which we can write the full diffeomorphism as

$$w(x) = \begin{cases} \frac{w_0^2}{w_0 + x} & x > 0 \\ w_0 + f^{-1}(-x) & x < 0 \end{cases} \tag{3.5}$$

for some $w_0 > 0$, where we have scaled the Poincaré coordinates $x, \bar{x}$ to set $f'(0) = 1$. In writing this, we have required that w and its first derivative are continuous at $x = y = 0$,

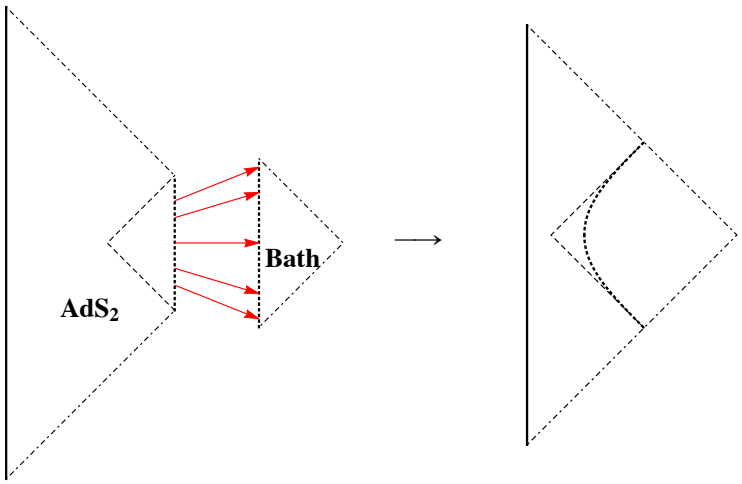

Figure 4. Penrose diagrams of the identification for the time-symmetric coupled system. Only the future-half of the diagram is relevant to the physical system in which the coupling is switched on only at $t = 0$. At $t = 0$ the bath is prepared in the Minkowski half-line vacuum. Dash-dotted lines denote future and past null infinity in the relevant Minkowski space. The remaining bath boundary (dotted) is identified with the right boundary of AdS_2 using the diffeomorphism f. We describe the state of the matter fields in the patch of AdS_2 shown, bounded by the dash-dotted lines denoting future and past Cauchy horizons. The dashed lines denote the event horizons of the right boundary. The resulting state is the half-line vacuum in the Minkowski half-space with auxiliary metric $dwd\bar{w}$, for which the Penrose diagram is shown on the right. In the physically relevant limit $w_0 \to 0$, the worldline of the joined boundaries is pushed to the left in the w coordinates, becoming nearly null.

which ensures that correlation functions of primary operators are continuous and that the stress tensor one-point function is physically sensible, as we will see in a moment. We have also made use of the symmetry under rescaling w to fix a free coefficient, and the extension of f to negative values as an odd function.

This map suffices for the piece of AdS_2 bounded by the Cauchy horizons at w or $\bar{w} \to \infty$ (x or $\bar{x} \to t_\infty := \lim_{u\to\infty} f(u)$), and by the Poincaré horizons $w = 0$ ($x \to \infty$) and $\bar{w} = 0$ ($\bar{x} \to \infty$). The map to w coordinates can be straightforwardly extended past the Poincaré horizons by using coordinates on AdS_2 that cover a larger patch.

We can now use this map and the anomaly (3.4) to compute the stress tensor

one-point function in the x and y frames (recalling that we have set $f'(0) = 1$):

$$\langle T_{xx}(x)\rangle = -\frac{c}{24\pi}\{w, x\} = E_S\delta(x) - \frac{c}{24\pi}\Theta(-x)\{y, x\} \tag{3.6}$$

$$\langle T_{yy}(y)\rangle = -\frac{c}{24\pi}\{w, y\} = E_S\delta(y) + \frac{c}{24\pi}\Theta(y)\{f(y), y\} \tag{3.7}$$

$$E_S = \frac{c}{24\pi}\left(\frac{2}{w_0} - f''(0)\right) \tag{3.8}$$

The δ-function contributions arise from the discontinuous second derivative of w, and at this point we are free to tune its coefficient E_S by choice of w_0.

However, if we use this result for finite E_S, the answers we will find are not compatible with the physics we are trying to capture. In particular, this will give nonzero connected two-point functions for an operator in AdS_2 and an operator in the bath, both spacelike separated from the point at which the coupling is turned on, though the states should be uncorrelated at such points. The resolution is that we must take $E_S \to \infty$. This is to be expected since there is of no finite energy state in a quantum field theory that is uncorrelated at finite separation. Without using an explicitly regulated model we must accept either acausal correlations or infinite energy. For example, using a lattice regulated theory where we join a pair of spin chains at time $t = 0$, energy will be introduced on the scale of the cutoff, $E_S \sim \varepsilon^{-1}$, and typically there will be faster than light propagation (since the relativistic theory only emerges in the infrared) that is negligible only on scales far longer than the cutoff. In the end, we will take E_S as some intermediate scale between AdS and Planck, the former to avoid dependence on details of the regularization, and the latter for validity of effective field theory, which we presume has a subPlanckian cutoff.

Taking $E_S \to \infty$ ($w_0 \to 0$), we find a simple limiting map to the upper half-plane:

$$w(x) \sim \begin{cases} \left(\frac{12\pi}{c}E_S\right)^{-2}\frac{1}{x} & x > 0 \\ f^{-1}(-x) & x < 0 \end{cases} \tag{3.9}$$

We have precisely the same map for right-moving coordinates $\bar{w}(\bar{x})$.

This approach gives an alternative route to previous results on local quantum quenches [40–42]. These analyses use an explicit regulating prescription, by offsetting the removal of the boundary slightly in Euclidean time. The results are equivalent after identifying E_S with the regulator.

To calculate correlation functions and entropies, we now only need to compute the half-plane correlators in the metric $dwd\bar{w}$, using the maps $w(x), \bar{w}(\bar{x})$, and then transform to the physical metric with the appropriate Weyl factor:

$$ds^2 = \Omega_w^{-2} dw d\bar{w}, \qquad \Omega_w = \frac{x + \bar{x}}{2}\sqrt{w'(x)\bar{w}'(\bar{x})} \tag{3.10}$$

3.3 Entropy in the half-plane

Since we have mapped the state of the system to the half-plane, we now need to compute the relevant quantities there, which for us is the entropy of a single interval. We here fix a definition for a renormalized entropy, and then review its computation for a single interval in the half-plane.

A convenient way to compute entropies of intervals in CFT_2 is to use the replica trick to compute the Rényi entropies $S^{(n)} = -\frac{1}{1-n}\operatorname{Tr}\rho^n$ for integer n, as path integrals on n copies of the geometry glued appropriately along the region in question, and take a formal $n \to 1$ limit to recover von Neumann entropy $S = -\operatorname{Tr}\rho\log\rho$. This is particularly powerful for two-dimensional CFTs [41] because the n copies of the geometry in the theory $\mathcal{C}$ can be described on a single copy of the geometry by correlation functions in an orbifold theory $\mathcal{C}^n/\mathbb{Z}_n$, where we take n copies of the original theory and quotient by the symmetry of cyclic permutation[4]. In this description, the boundary conditions for the replica manifold are implemented by inserting twist operators $\sigma, \tilde{\sigma}$ at the left and right endpoints of the interval in question:

$$S^{(n)} = -\frac{1}{n-1}\log\langle\sigma(x_1,\bar{x}_2)\tilde{\sigma}(x_2,\bar{x}_2)\rangle_{\mathcal{C}^n/\mathbb{Z}_n} \tag{3.11}$$

We now use the powerful fact that the twist operators are local, primary operators in the orbifold theory, with dimension

$$\Delta_n = \frac{c}{12}\frac{(n-1)(n+1)}{n}\,. \tag{3.12}$$

Since it will be important for us to work on curved manifolds, we emphasize that this property determines correlation functions on a manifold with Weyl rescaled metric $\Omega^{-2}g$ in terms of those with metric g (the expectation values here are normalized by the partition function without operator insertions, so the conformal anomaly cancels):

$$\langle\sigma(x_1,\bar{x}_2)\tilde{\sigma}(x_2,\bar{x}_2)\rangle_{\Omega^{-2}g} = \Omega(x_1,\bar{x_1})^{\Delta_n}\Omega(x_2,\bar{x_2})^{\Delta_n}\langle\sigma(x_1,\bar{x}_2)\tilde{\sigma}(x_2,\bar{x}_2)\rangle_g \tag{3.13}$$

The replica manifold is singular at the endpoints of the interval, which leads to a divergent Rényi entropy; any local regulator of this divergence can be simply absorbed into the normalization of the twist operators. We can theorefore define renormalized Rényi entropies, and hence von Neumann entropies, by choosing a canonical normalization (which implicitly introduces an infrared length scale, for us the AdS scale) where the identity appears in the OPE with coefficient one:

$$\sigma(x_1)\tilde{\sigma}(x_2) \sim |x_1 - x_2|^{-2\Delta_n} \quad \text{as } x_2 \to x_1 \tag{3.14}$$

[4]This is not strictly true in all states, since the orbifold theory correlators may involve an unwanted sum over twisted sectors, but it will hold for all the examples we use.

Due to the primary transformation property, this normalization holds independently of the metric, when $|x_1 - x_2|$ denotes the proper distance between the operator insertions. In particular, with this normalization, the renormalized entropy of a single interval of length ℓ on the line in vacuum is $S = \frac{c}{3}\log\ell$.

Taking the $n \to 1$ limit of (3.13), we find how the von Neumann entropy behaves under Weyl transformations:

$$S_{\Omega^{-2}g} = S_g - \frac{c}{6}\sum_{\text{endpoints}} \log\Omega \tag{3.15}$$

This can be thought of as arising from a local rescaling of the cutoff scale at each endpoint of the interval.

We can now address the pertinent example of the entropy of a single interval on the half-plane. We begin with the case when the interval contains the boundary, which will be used to compute the entropy in AdS before turning on the coupling, and also for the total entropy of system and bath combined. For this, we insert only one twist operator in the bulk, so we need only write down the most general conformally invariant one-point function on the half-plane:

$$\langle\sigma(w,\bar{w})\rangle_{\text{UHP}} = \frac{g_n}{(w+\bar{w})^{\Delta_n}} \tag{3.16}$$

We could think of there being also a twist operator inserted on the boundary, but boundary twist operators are topological: they have zero scaling dimension, and the insertion point can be freely deformed without changing the correlation functions. We note also that the interval of interest could lie on either side of σ and give the same result, reproducing the expectation from purity of the total state.

Taking the $n \to 1$ limit, we find

$$S = \frac{c}{6}\log(w+\bar{w}) + \log g, \tag{3.17}$$

where $\log g = -\partial_n \log g_n|_{n=1}$ is the Affleck-Ludwig boundary entropy [43].

Next, we take the interval to have both endpoints in the bulk of the system, away from the boundary. Now, since we are on the half-plane, for two points we can construct a conformally invariant cross ratio (invariant under the $PSL(2,\mathbb{R})$ which fixes the boundary on the imaginary axis):

$$\eta = \frac{(w_1+\bar{w}_1)(w_2+\bar{w}_2)}{(w_1+\bar{w}_2)(w_2+\bar{w}_1)} \tag{3.18}$$

The two-point function of twist-operators does not have a fixed functional form, but contains an undetermined function $G_n(\eta)$, which can depend on the theory and bound-

ary conditions:

$$\langle \sigma(w_1,\bar{w}_1)\tilde{\sigma}(w_2,\bar{w}_2)\rangle_{\text{UHP}} = \frac{G_n(\eta)}{((w_1-w_2)(\bar{w}_1-\bar{w}_2)\eta)^{\Delta_n}} \tag{3.19}$$

In the kinematics relevant for entropy, where the endpoints of the interval remain spacelike separated, the cross-ratio η varies between one and zero. Approaching either of these limits, G_n is determined by either a bulk OPE or boundary operator expansion:

$$\eta \to 1, \qquad G_n(\eta) \to 1 \qquad (\sigma\tilde{\sigma} \text{ OPE limit}) \tag{3.20}$$

$$\eta \to 0, \qquad G_n(\eta) \to g_n^2 \qquad (\text{boundary limit}) \tag{3.21}$$

Taking the $n \to 1$ limit, we find the von Neumann entropy

$$S = \frac{c}{6}\log\left[(w_1-w_2)(\bar{w}_1-\bar{w}_2)\eta\right] + \log G(\eta), \tag{3.22}$$

where $\log G(\eta) = -\partial_n \log G_n(\eta)|_{n=1}$, satisfying $G(1) = 1$, $G(0) = g^2$.

3.4 Mapping to AdS_2

With all the ingredients in place, it remains only to put them together and compute various quantities of interest in AdS.

First, we compute the expectation value of the stress tensor in AdS_2. In the x coordinate in the flat $dxd\bar{x}$ metric this is given by (3.6), and there is no anomaly from the Weyl factor to transform to the physical Poincaré AdS metric (3.1):

$$\langle T_{--}(x^-)\rangle_{\text{AdS}_2} = E_S\delta(x^-) - \frac{c}{24\pi}\{y^-,x^-\}\Theta(x^-), \tag{3.23}$$

where $y^- = f^{-1}(x^-)$. Note that $\{y^-,x^-\}$ will typically be positive, so this represents an injection of negative energy into AdS (in the Hartle-Hawking state on AdS, by $\mathfrak{sl}(2)$ invariance the stress tensor expectation value vanishes, aside from the trace which is identically a constant determined by the curvature and anomaly). We can similarly construct $\langle T_{++}\rangle$ by time-reversal invariance, but for the positive times we are interested in it vanishes. This result is used to determine the dynamics of the reparameterization mode f in section 2.2.

Next, we compute the entropies of an interval in AdS_2, along with the bath. This is equal to the entropy of the purifying system, so can be used to verify that the quantum extremal surface of the purifier remains at the bifurcate horizon for all times.

For this, we use the result (3.17), along with the transformation (3.15). The result is simplest when the boundary of the interval is spacelike separated from the coupling,

$x^+ > 0$ and $x^- < 0$. In that case, we get the same answer as in pure AdS as required by causality, which is a constant by $SL(2,\mathbb{R})$ invariance of the spacetime and state:

$$S = \frac{c}{6}\log 2 + \log g \qquad (x^+ > 0, \quad x^- < 0) \tag{3.24}$$

For a more nontrivial result, we take the endpoint to lie in the future of the shock, $x^+ > x^- > 0$:

$$S = \frac{c}{6}\log\left(\frac{24\pi E_S}{c}\frac{x^+ y^- \sqrt{f'(y^-)}}{x^+ - x^-}\right) + \log g, \quad x^+ > x^- = f(y^-) > 0 \tag{3.25}$$

An example of this is to compute the total entropy of the bath at physical time u or Poincaré time $t = f(u) = \frac{x^+ + x^-}{2}$, where we cutoff AdS_2 at $z = \frac{x^+ - x^-}{2} = \epsilon f'(u)$:

$$S = \frac{c}{6}\log\left(\frac{12\pi E_S}{\epsilon c}\frac{u f(u)}{\sqrt{f'(u)}}\right) + \log g \tag{3.26}$$

Finally, we compute the entropy of an interval in AdS_2 using (3.22). There are two main cases of interest, with either one or both endpoints to the future of the shock. We begin with the case with one endpoint to the future, $x_1^\pm > 0$, and one endpoint to the past, $x_2^+ > 0, x_2^- < 0$. In this case, the cross-ratio is nontrivial,

$$\eta = \frac{x_1^+(x_2^+ - x_2^-)}{x_2^+(x_1^+ - x_2^-)} \tag{3.27}$$

and the entropy is

$$S = \frac{c}{6}\log\left[\frac{48\pi E_S}{c}\frac{-y_1^- x_1^+ x_2^-(x_2^+ - x_1^+)\sqrt{f'(y_1^-)}}{x_2^+(x_1^+ - x_1^-)(x_1^+ - x_2^-)}\right] + \log G(\eta). \tag{3.28}$$

If we take the whole interval to lie to the future of the shock, $x_\pm^{(1,2)} > 0$, the cross ratio goes to $\eta = 1$, and we have

$$S = \frac{c}{6}\log\left[\frac{4(y_1^- - y_2^-)(x_2^+ - x_1^+)\sqrt{f'(y_1^-)f'(y_2^-)}}{(x_1^+ - x_1^-)(x_2^+ - x_2^-)}\right]. \tag{3.29}$$

As a consistency check, we note that all entropies are invariant under the residual symmetries of AdS_2, namely the $PSL(2,\mathbb{R})$ transformations $x \mapsto \frac{ax}{cx+d}$ that fix zero. This means that we act on all coordinates labelled by x, and on f, but not on y.

Finally, take the limit of the above answers where one endpoint goes to the AdS boundary at physical time u, Poincaré time $t = f(u)$, regulated to lie on the cutoff surface $z = \epsilon f'(u)$:

$$S = \begin{cases} \frac{c}{6} \log \left[\frac{24\pi E_S}{\epsilon c} \frac{-utx^-(x^+-t)}{x^+(t-x^-)\sqrt{f'(u)}} \right] + \log G \left(\frac{t(x^+-x^-)}{x^+(t-x^-)} \right) & x^- < 0 < t < x^+ \\ \frac{c}{6} \log \left[\frac{2(u-y^-)(x^+-t)}{\epsilon(x^+-x^-)} \sqrt{\frac{f'(y^-)}{f'(u)}} \right] & 0 < x^- < t < x^+ \end{cases} \tag{3.30}$$

This will be our main result required for finding quantum extremal surfaces.

3.5 Interpretation of entropies

Our formulas for the entropies admit simple quasiparticle interpretations in terms of freely propagating independent left- and right-moving degrees of freedom.

In (3.29), factor $(x_2^+ - x_1^+)$ gives the entropy of outgoing modes in the Poincaré vacuum. The factor $(y_1^- - y_2^-)$ computes the entropy of modes moving in from the bath, which are in the vacuum associated to the flat metric of the y coordinates. The remaining terms are conformal factors to transform to the physical AdS_2 metric, quantifying short-distance correlations. This can be thought of as adapting the cutoff to correspond to a local proper distance.

The case (3.28) when the interval straddles the shock allows for a similar interpretation. The outgoing modes are the same as for the previous case of (3.29). The factor y_1^- gives the entropy of the infalling state from the bath, starting only at time zero once the coupling to the bath is turned on. For the rest of the interval, the ingoing modes have been reflected off the boundary, so their entropy includes contributions from entanglement with outgoing modes, which is quantified using (3.22). The constant factor, proportional to $\log E_S$, represents the entropy of the shock itself, which is entangled with the corresponding shock propagating into the bath.

4 The quantum extremal surfaces

Now that we have computed the entropy of matter fields and solved the gravitational dynamics, we have all the required ingredients to locate the quantum extremal surface, and hence the entanglement wedge as the black hole evaporates.

4.1 The generalized entropy

A quantum extremal surface is defined to extremize the generalized entropy, which is the sum of area in Planck units and the entropy of bulk matter. However, on first

sight this is ambiguous, because both ingredients depend on choices of regulator, or the energy scale of the bulk effective field theory we choose to use. The resolution is that the generalized entropy is in fact better defined than its constituents (see [44] for a recent discussion), which can be seen very explicitly in our model.

First, we can define the generalized entropy in terms of renormalized, or infrared quantities:

$$S_{\text{gen}} = \frac{\phi_0^{(\text{Ren.})} + \phi^{(\text{Ren.})}}{4G_N} + S^{(\text{Ren.})} \tag{4.1}$$

For example, we can define the renormalized entropy $S^{(\text{Ren.})}$ as a finite quantity using the scheme of section 3. This implicitly involved introducing a finite, infrared scale, in our case the AdS scale.

For an alternative definition, we can use the 'bare' quantities, explicitly cutting off the bulk matter theory at some UV scale ε. For our calculations, we treat the gravitational sector classically, but integrating out the matter sector will nonetheless renormalize the gravitational parameters.

Even though our matter sector is conformal, regularization must introduce a length scale. Explicitly, the CFT partition function on some space with Euler character χ depends on the typical length scale L (for us, L is the AdS scale):

$$\mathcal{Z}_{\text{CFT}}[L] \propto \left(\frac{L}{\varepsilon}\right)^{\frac{c}{6}\chi} \tag{4.2}$$

The functional dependence on L is determined by the trace anomaly, but it necessitates the introduction of the length scale ε. Note that this has the same form as the topological term in the action:

$$-\frac{\phi_0}{16\pi G_N}\left[\int \mathcal{R} + 2\int_\partial K\right] = -\frac{\phi_0}{4G_N}\chi \tag{4.3}$$

We can therefore absorb this scale dependent normalization of the partition function by an additive renormalization of the dilaton.

$$-\frac{\phi_0^{(\text{Bare})}}{4G_N}\chi - \log \mathcal{Z}_{\text{CFT}} = -\frac{\phi_0^{(\text{Ren.})}}{4G_N}\chi, \qquad \phi_0^{(\text{Ren.})} = \phi_0^{(\text{Bare})} - \frac{2}{3}c\, G_N \log \varepsilon \tag{4.4}$$

If we use the same cutoff scale for the entropies, the bare and renormalized quantities are related by a similar shift

$$S^{(\text{Bare})} = S^{(\text{Ren.})} - N\frac{c}{6}\log \varepsilon, \tag{4.5}$$

where N is the number of endpoints of an interval in question. Combining these two results, we see explicitly that we can use bare quantities in the definition of the generalized entropy, and find a finite, regulator independent result that matches the one obtained using renormalized, infrared quantities.

4.2 Early times

Before we have disturbed the black hole by coupling to an external system, the quantum extremal surface coincides with the classical extremal surface, at the bifurcation point of the original black hole horizon. For a time after the coupling is turned on, the QES does not stray far; we begin by finding its location at these early times.

The bifurcation surface is of course causally disconnected from the process of coupling to the auxiliary system, so the nearby geometry is unaffected, but the quantum extremal surface can nonetheless begin to move as soon as the evaporation process begins. The relevant effect is that the fields near the bifurcation surface are entangled with the first few Hawking quanta to escape; an entangling surface closer to the boundary captures less of this entanglement, and hence has lower entropy. Moreover, this entanglement changes linearly with distance, whereas the classical area term varies quadratically, with the result that the quantum extremal surface moves out towards the boundary in a spacelike direction, of order a Planck distance if evaporation proceeds for a thermal time.

We now show this quantitatively in our model. For a surface located in the $x^- < 0$ region before the shock, the area is simply that of the original black hole, while the entropy is computed from the first case in (3.30).

$$\begin{aligned} S_{\text{gen}} &= \frac{\phi}{4G_N} + S, \qquad \phi = \phi_0 + 2\bar{\phi}_r \frac{1-(\pi T_0)^2 x^+ x^-}{x^+ - x^-}, \\ S &= 2k\frac{\bar{\phi}_r}{4G_N}\left[\log\left(\frac{24\pi E_S}{\epsilon c}\frac{-x^-(x^+-t)tu}{x^+(t-x^-)\sqrt{f'(u)}}\right) + \mathcal{G}\left(\frac{t(x^+-x^-)}{x^+(t-x^-)}\right)\right] \end{aligned} \tag{4.6}$$

We have here defined $\mathcal{G}(\eta) = \frac{6}{c}\log G(\eta)$ for brevity of notation; in particular $\mathcal{G}(1) = 0$.

We now simply make the ansatz that the deviation from the horizon is small, taking $x^\pm \mp \frac{1}{\pi T_0}$ of order k, and solve for stationarity of S_{gen} in the $k \to 0$ limit.

$$\frac{1}{\pi T_0} - x^+ \sim \frac{k}{(\pi T_0)^2}(\eta - \eta(1-\eta)\mathcal{G}'(\eta)) \tag{4.7}$$

$$x^- + \frac{1}{\pi T_0} \sim \frac{k}{(\pi T_0)^2}\left(\frac{\eta}{1-\eta} - \eta\mathcal{G}'(\eta)\right) \tag{4.8}$$

$$\eta = \frac{2\pi T_0 t}{1+\pi T_0 t} \tag{4.9}$$

This solution describes a QES that starts at the bifurcation surface at $t = 0$, moves outwards in a spacelike direction towards the boundary, before bending to an almost null path that runs along the horizon, approaching constant outgoing coordinate x^+.

When parameterized in terms of the Poincaré time t on the boundary, this family of surfaces moves steadily along the horizon, and can be continued far away from the

bifurcation surface, beyond the point where our approximation breaks down. However, in terms of the physical time u, they end up settling at final location, because as $u \to \infty$, t approaches a finite limit $t_\infty \sim \frac{1}{\pi T_1} < \frac{1}{\pi T_0}$. In the limit where the shock from turning on the coupling adds much more than a thermal unit of energy ($E_S \gg T_0$, or equivalently $T_1 - T_0 \gg k$), this location is sufficiently close to the bifurcation surface to be within the regime of our approximation:

$$1 - \eta \gtrsim \frac{T_1 - T_0}{T_1 + T_0} \tag{4.10}$$

For sufficiently early times, we have $1-\eta \sim e^{-2\pi T_1 u}$, until η settles down to its maximum for $u \gtrsim -\frac{1}{T_1} \log\left(\frac{T_1-T_0}{T_1+T_0}\right)$.

We can now evaluate the generalized entropy on the extremal surface, to track the fine-grained entropy of the black hole (this equation valid for times $u \ll k^{-1}$):

$$S_{\text{gen}} \sim \frac{\phi_0 + 2\pi T_0 \bar{\phi}_r}{4G_N} + \frac{c}{6} \log\left(\frac{24\pi E_S}{\epsilon c} \frac{u \sinh(\pi T_1 u)}{\pi T_1}(1-\eta)\right) + \frac{c}{6}\mathcal{G}(\eta) \tag{4.11}$$

At vey early times, we have a large logarithmic increase in the entropy, due to production of short-distance entanglement with the bath; we should take this seriously only for times larger than the regulator scales ϵ and E_S^{-1}, so the entropy here increases from the equilibrium value. Next, the $\log(1-\eta)$ term takes over, giving a linear decrease in entropy as thermal Hawking radiation escapes into the bath. Eventually, after a scrambling time

$$t_{\text{min } S} = \frac{1}{2\pi T_1} \log\left(\frac{T_1}{T_1 - T_0}\right) = \frac{\beta_1}{2\pi} \log\left(\frac{E(T_1)}{E_S}\right) \tag{4.12}$$

for $E(T_1)$ the energy of a black hole at temperature T_1 and $T_1 - T_0 \ll T_0$, η settles down to its maximum, entropy reaches a local minimum, and then entropy begins to increase linearly. The system begins to access all the extra states opened up by the addition of energy, so we have an increase in thermal entropy purified by the outgoing Hawking radiation, at rate $\frac{c}{6}\pi T_1$. The black hole is now 'young', in the sense that its fine grained entropy is less than the thermodynamic entropy at same energy. This linear increase is the first part of the familiar Page curve [45], though for an old black hole after a perturbation rather than for a hole formed recently from collapse. We note that the time (4.12) is precisely of the form anticipated in section 1.1 with coefficient $\alpha_S = 1$.

The surfaces described here are the only relevant solutions to the quantum extremization equations for $x^- < 0$. There exists another family of surfaces at small negative x^- which extremize S_{gen}, but these always have much larger generalized entropy.

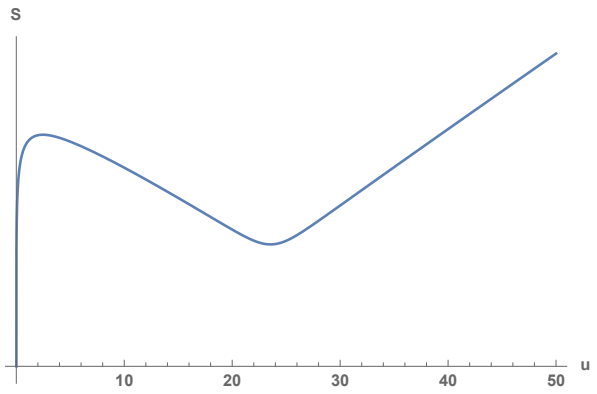

Figure 5. The von Neumann entropy of the black hole, as computed from the generalized entropy on the quantum extremal surface, at early times after coupling to the bath. In this plot, we have $T_0 = \frac{1}{2\pi}$ and $\frac{T_1 - T_0}{T_0} = 10^{-10}$.

4.3 A QES inside the shock

The next place for a quantum extremal surface to exist is inside the shock itself, at x^- very close to zero. While we do not have a microscopic model for the entropy within the shock itself, these details will not matter, and we can focus on the generalized entropy for small negative or positive x^-, just before or after the shock falls in.

Expanding the entropy near the shock, we have the following:

$$S \sim \frac{c}{6} \log \left[\frac{2u(x^+ - t)}{\epsilon x^+ \sqrt{f'(u)}} \right] + \frac{c}{6} \begin{cases} \log \left(\frac{12\pi E_S}{c} \frac{-x^-}{1} \right) & x^- < 0 \\ \left(\frac{1}{x^+} - \frac{1}{u} \right) x^- & x^- > 0 \end{cases} \tag{4.13}$$

We should take the first line seriously only for $x^- \gg E_S^{-1}$ (which can be very close to the shock because we are taking $E_S \gg 1$). For an interval that straddles the shock, there is a large, $\log E_S$ contribution to the entropy, from entanglement between the shock falling into the black hole and its counterpart propagating into the auxiliary system. The entropy drops off rapidly as we approach $x^- = 0$, since the initial product state between system and bath carries no short-distance entanglement.

To look for a quantum extremal surface, we first extremize the generalized entropy in the ∂^+ direction, using the dilaton $\phi - \phi_0 = \frac{2\bar{\phi}_r}{x^+}$:

$$\partial_+ S_{\text{gen}} = 0 \implies x^+ = \frac{t}{1 - kt} \tag{4.14}$$

Since the dilaton is decreasing as a function of x^+, we need to balance this against a rapid increase from the entropy. This comes from the lightcone singularity as x^+

approaches t, a phenomenon we will revisit later. At this location, the generalized entropy is maximal for variations parallel to the shock.

For a quantum extremal surface to exist on the shock, we will also require that the generalized entropy is locally minimal under variations in the outgoing, x^- direction:

$$\frac{4G_N}{\bar{\phi}_r}\partial_- S_{\text{gen}} \sim \begin{cases} 2\frac{1}{(x^+)^2} - 2(\pi T_0)^2 + \frac{2k}{x^-} & x^- < 0 \\ 2\frac{1}{(x^+)^2} - 2(\pi T_1)^2 + \frac{2k}{x^+} - \frac{2k}{u} & x^- > 0 \end{cases} \tag{4.15}$$

It appears that the singularity in the entropy will cause this to be decreasing just before the shock, but generically this will only be large enough to overcome the large dilaton increase within the microscopic scale of the shock E_S^{-1}, so the result should not be trusted. However, this is no longer true very close to the horizon, for $\frac{1}{\pi T_0} - x^+ \ll \frac{kE_S}{T_0^3}$, in which case we can have a quantum extremal surface on the condition that the entropy begins to increase again for $x^- > 0$. This is satisfied as long as

$$x^+ - \frac{1}{\pi T_1} \lesssim \frac{k}{2(\pi T_1)^2} \implies t - \frac{1}{\pi T_1} \lesssim -\frac{k}{2(\pi T_1)^2} \tag{4.16}$$

Translating this into the boundary proper time, using the expression (2.27) for t expanded to first order in small ku, and the expression (2.24) for t_∞, we find that there is a candidate quantum extremal surface supported on the shock for times

$$u < \frac{1}{2\pi T_1} \log\left(\frac{8\pi T_1}{3k}\right). \tag{4.17}$$

The generalised entropy for these surfaces is larger at any given boundary time than the corresponding entropy (4.11) for the extremal surface we found in the previous section. This is simply because the area has been increased by the energy from the shock, and the additional entropy is far too small to make up for it. These surfaces can never have minimal S_{gen}, so are not relevant for determining the entanglement wedge.

4.4 Soon after the shock

We now look for quantum extremal surfaces in the region $x^- > 0$, to the future of the coupling to the bath, and the resulting shock of energy. At this point, if one thinks of a quantum extremal surface as a small perturbation to a classical extremal surface, it may seem that none should exist in this regime. In particular, for any $x^- > 0$, the area is never close to being stationary under variations in the + direction.

Nonetheless, the bulk entropy can compete with the area to introduce an extremum of the generalized entropy which is far from any classical extremal surface. The necessary enhancement of the variation of entropy comes from the large amount of entanglement in the bulk fields at short distances, in particular in the outgoing Hawking

radiation. Consider the generalized entropy of a surface at some fixed x^- as we decrease x^+, moving the surface out towards the boundary in a past null direction. The area steadily increases all the way, but the entropy at some fixed boundary time decreases, from the loss of entanglement with the previous Hawking radiation. When the surface approaches null separation from the boundary time of interest, the rate of decrease of entropy is sufficient to overcome the increase in area, so the generalized entropy has a maximum in this direction. This will typically happen close to the apparent horizon, in which case the variation of area along the horizon is in any case small. It is then unsurprising that the area can also be extremized in the direction parallel to the horizon, leading to our quantum extremal surface.

To see this more quantitatively, take first the case when x^- is of order one. At this early time, for many purposes we can ignore the backreaction resulting from the slow leakage of energy into the bath, so can take the dilaton profile of the static black hole solution with temperature T_1 (including the effect of the shock). For extremizing in the x^+ direction, the only relevant piece of the bulk entropy is proportional to $\log(x^+ - t)$, quantifying the entanglement of outgoing modes. Keeping only these terms, we find

$$\begin{aligned} \frac{4G_N}{\phi_r}\partial_+ S_{\text{gen}} &\sim -2\frac{1-(\pi T_1 x^-)^2}{(x^+ - x^-)^2} + \frac{2k}{x^+ - t} \\ \implies x^+ - t &\sim \frac{k}{(\pi T_1)^2}\frac{1-\pi T_1 x^-}{1+\pi T_1 x^-}, \end{aligned} \tag{4.18}$$

where in the last equation we have also approximated $t, x^+ \sim \frac{1}{\pi T_1}$, which is required for the quantum extremal surface to exist in the $x^- > 0$ region.

Now it remains only to extremize in the x^- direction, along the horizon. Close to the horizon, the variation of the background dilaton is suppressed (proportional to $x^+ - \frac{1}{\pi T_1}$), so we must also include the backreaction:

$$\begin{aligned} (x^+ - x^-)^2\partial_-\left(\frac{I}{x^+ - x^-}\right) &\sim \int_0^{x^-} dt (x^+ - t)^2 \left\{\frac{\tanh^{-1}(\pi t T)}{\pi T}, t\right\} \\ &\sim \frac{2x^-}{1+\pi T_1 x^-} \end{aligned} \tag{4.19}$$

Combining this with the unbackreacted term, where we use (4.18) to determine the deviation of x^+ from the horizon, we find

$$\partial_- \phi \sim \bar{\phi}_r \left(\frac{(2\pi T_1)^2(1-\pi T_1 x^+)}{(1-\pi T_1 x^-)^2} + \frac{2(\pi T_1)^2 k x^-}{(1+\pi T_1 x^-)(1-\pi T_1 x^-)^2}\right) \tag{4.20}$$

Finally, we need the variation of the entropy. We use the early time, unbackreacted solution of the reparameterization, $f(u) \sim \frac{1}{\pi T_1}\tanh(\pi T_1 u)$, so in particular we have

$f'(y^-) \sim 1 - (\pi T_1 x^-)^2$. For times $u \gg 1$, the variation of the $\log(u - y^-)$ term is negligible.

$$\partial_- S = \frac{c}{6} \frac{\pi T_1}{1 - (\pi T_1 x^-)^2} \tag{4.21}$$

Putting this together we find

$$\frac{4G_N}{\bar{\phi}_r} \partial_- S_{\text{gen}} \sim \frac{(2\pi T_1)^2}{(1 - \pi T_1 x^-)^2} \left[(1 - \pi T_1 x^+) + \frac{k}{2\pi T_1} \frac{1}{1 + \pi T_1 x^-} \right], \tag{4.22}$$

which gives us our condition for quantum extremality in the $-$ direction:

$$\partial_- S_{\text{gen}} = 0 \implies x^+ - \frac{1}{\pi T_1} \sim \frac{k}{2(\pi T_1)^2} \frac{1}{\pi T_1 x^- + 1} \tag{4.23}$$

This locates our quantum extremal surface, which we now write in terms of proper time u. In the relevant regime, with $ke^{2\pi T_1 u}$ of order one, this is related to Poincaré time by $t - \frac{1}{\pi T_1} \sim \frac{k}{(2\pi T_1)^2} - \frac{2}{\pi T_1} e^{-2\pi T_1 u}$ (using (2.27), and requiring the expansion (2.24) of t_∞).

$$x^- \sim \frac{1}{\pi T_1} \frac{3ke^{2\pi T_1 u} - 8\pi T_1}{3ke^{2\pi T_1 u} + 8\pi T_1} \tag{4.24}$$

$$t_\infty - x^+ \sim -\frac{2}{3} t_\infty e^{-2\pi T_1 u} \tag{4.25}$$

For this to be valid, we require that the extremum lies to the future of the shock in the $x^- > 0$ region, so $u > \frac{1}{2\pi T_1} \log\left(\frac{8\pi T_1}{3k}\right)$. Before this time, this family of surfaces joins continuously onto those of the previous section, which live somewhere inside the shock.

4.5 Later times

The quantum extremal surfaces found so far have relied on approximations to the dilaton profile which are valid only at early times, and our result holds only when $e^{-2\pi T u} \gg k^2$. Going to higher orders in the small parameter k does not help much: it would extend validity to times when $e^{-2\pi T u}$ scales as a higher power in k, which accesses only a few scrambling times. To go to times of order k^{-1}, when a significant fraction of the black hole has evaporated, we require a different approach to treat the dilaton. The idea is to determine the dilaton by working backwards from very late times, and using the fact that it vanishes on the boundary at the Poincaré time $t = t_\infty$, corresponding to $u \to \infty$ in proper time. This prevents the exponential growth of errors in the approximations we make.

We continue the same intuition that the lightcone singularity balances the area in the ingoing direction, making the ansatz $t_\infty - x^+ \ll t_\infty - x^-$ to put the surface close

to the horizon. Begin with the variation of the dilaton in the plus direction, using the integral expression (2.32) for the effect of backreaction:

$$(x^+ - x^-)^2 \partial_+ \phi = \bar{\phi}_r \left[2(\pi T_1 x^-)^2 - 2 - k \int_0^{x^-} dt (x^- - t)^2 \{u, t\} \right] \tag{4.26}$$

The right hand side is a function of x^- only, which must be negative for any $x^- < t_\infty$, since ϕ goes to infinity at the boundary. At $x^- = t_\infty$, it vanishes, so in that limit the integral must give

$$I_\infty = \int_0^{t_\infty} dt (t_\infty - t)^2 \{u, t\} = \frac{2}{k}((\pi T_1 t_\infty)^2 - 1). \tag{4.27}$$

We can now work backwards from this result to estimate the integral fo x^- in the range of interest, corresponding to ingoing time y^- of order k^{-1}:

$$\begin{aligned} \partial_- \int_0^{x^-} dt (x^- - t)^2 \{u, t\} &= 2 \int_0^{x^-} dt (x^- - t) \{u, t\} \\ &\sim \int^{x^-} \frac{x^- - t}{(t_\infty - t)^2} dt \\ &\sim -\log\left(\frac{t_\infty - x^-}{t_\infty} \right) + O(1) \\ \implies \int_0^{x^-} dt (x^- - t)^2 \{u, t\} &\sim \tfrac{2}{k}((\pi T_1 t_\infty)^2 - 1) + (t_\infty - x^-) \log\left(\frac{t_\infty - x^-}{t_\infty} \right) \end{aligned}$$

Here we use a late time approximation for $\{u, t\}$, for which the peak of the integrand at $t_\infty - t = 2(t_\infty - x^-)$ gives the logarithm.

We can now assemble the pieces to give the desired variation of the dilaton:

$$\partial_+ \phi \sim -\bar{\phi}_r \left[(2\pi T_1)^2 \frac{t_\infty}{t_\infty - x^-} - \frac{k}{t_\infty - x^-} \log\left(\frac{t_\infty}{t_\infty - x^-} \right) \right] \tag{4.28}$$

$$\sim -4\pi T_1 \bar{\phi}_r \frac{e^{-\frac{k}{2} y^-}}{t_\infty - x^-} \tag{4.29}$$

In the second line, we have used the approximate form (2.27) for the reparameterization to substitute $x^- = f(y^-)$.

To extremize the generalized entropy, we balance this against the lightcone singularity in the entropy as before:

$$4\pi T_1 \frac{e^{-\frac{k}{2} y^-}}{t_\infty - x^-} \sim \frac{2k}{x^+ - t} \tag{4.30}$$

Now look at the variation of the dilaton parallel to the horizon:

$$(x^+ - x^-)^2 \partial_- \phi = \bar{\phi}_r \left[2 - 2(\pi T_1 x^+)^2 + k \int_0^{x^-} dt (x^+ - t)^2 \{u, t\} \right] \tag{4.31}$$

To analyse the integral, we expand in powers of $t_\infty - x^+$ and use the same approximation for $\{u, t\}$ as before:

$$\begin{aligned}
&\int_0^{x^-} dt (x^+ - t)^2 \{u, t\} - I_\infty \\
&\qquad = (t_\infty - x^+)^2 \int_0^{x^-} dt \{u, t\} - 2(t_\infty - x^+) \int_0^{x^-} dt (t_\infty - t) \{u, t\} \\
&\qquad\qquad - \int_{x^-}^{t_\infty} dt (t_\infty - t)^2 \{u, t\} \\
&\qquad \sim \frac{(t_\infty - x^+)^2}{2(t_\infty - x^-)} + (t_\infty - x^+) \log \left(\frac{t_\infty - x^-}{t_\infty} \right) - \tfrac{1}{2}(t_\infty - x^-)
\end{aligned}$$

This matches a result from early times. We can neglect the term quadratic in $t_\infty - x^+$, but the others can be relevant, as seen in the variation of the dilaton:

$$\begin{aligned}
\partial_- \phi &\sim \frac{\bar{\phi}_r}{(t_\infty - x^-)^2} \left[4\pi T_1 (t_\infty - x^+) + k (t_\infty - x^+) \log \left(\frac{t_\infty - x^-}{t_\infty} \right) - \tfrac{k}{2}(t_\infty - x^-) \right] \\
&\sim \frac{\bar{\phi}_r}{(t_\infty - x^-)^2} \left[4\pi T_1 (t_\infty - x^+) e^{-k y^-/2} - \tfrac{k}{2}(t_\infty - x^-) \right]
\end{aligned}$$

Finally, we compute the variation of the entropy. The $u - y^-$ term is unimportant, but we must include the $\sqrt{f'(y^-)}$ term:

$$\partial_{x^-} \log f'(y^-) \sim -\frac{1}{t_\infty - x^-} \tag{4.32}$$

From this, we get the variation of the bulk entropy,

$$\partial_- S \sim \frac{c}{12} \frac{1}{t_\infty - x^-}, \tag{4.33}$$

to find

$$\frac{4G_N}{\bar{\phi}_r} \partial_- S_{\text{gen}} = \frac{1}{(t_\infty - x^-)^2} \left[4\pi T_1 (t_\infty - x^+) e^{-k y^-/2} + \tfrac{k}{2}(t_\infty - x^-) \right], \tag{4.34}$$

and hence

$$x^+ - t_\infty \sim \frac{k}{8\pi T_1} (t_\infty - x^-) e^{k y^-/2}. \tag{4.35}$$

The important term in the variation of the bulk entropy appearing here has a simple interpretation that becomes clearer when rewritten in terms of the variation with respect to y^-:

$$\frac{\partial S}{\partial y^-} \sim -\frac{c}{12} 2\pi T_1 e^{-ky^-/2}$$

Over periods of time small compared to k^{-1}, for which the black hole can be regarded as almost static, this variation is approximately constant, and in fact equal to minus the (left-moving half of the) entropy density of the matter CFT at the temperature $2\pi T_1 e^{-ky^-/2}$ of the black hole at the relevant time. This entropy can be thought of as arising from the absence of the ingoing thermal matter that would be present in the Hartle-Hawking state (for which the entropy is constant by $SL(2)$ invariance), but has instead escaped into the bath.

Putting this together with extremization in the x^+ direction, we can locate the quantum extremal surface:

$$x^+ - t_\infty \sim \frac{t_\infty - t}{3} \tag{4.36}$$

$$t_\infty - t \sim \frac{3k}{8} t_\infty e^{\frac{k}{2}y^-} (t_\infty - x^-) \tag{4.37}$$

At this ingoing time x^-, we have

$$t_\infty - x^+ \sim \begin{cases} -\frac{1}{3}(t_\infty - t) & \text{(QES)} \\ 0 & \text{(event horizon)} \\ \frac{1}{3}(t_\infty - t) & \text{(apparent horizon } \partial_- \phi = 0) \end{cases} \tag{4.38}$$

so the QES is the same distance inside the event horizon as the apparent horizon is outside.

Expressing this entirely in terms of the boundary proper time u and corresponding ingoing coordinate y^-, we have our final location for the quantum extremal surface:

$$u \sim y^- + \frac{e^{\frac{k}{2}y^-}}{2\pi T_1} \log\left(\frac{8\pi T_1 e^{-\frac{k}{2}y^-}}{3k}\right) \tag{4.39}$$

In the case that this is the quantum extremal surface of minimal generalized entropy (of which more in a moment), this gives a beautiful quantum geometric realization of the Hayden-Preskill experiment. The quantum extremal surface demarks the region of spacetime to which the system has access. This region is bounded to the future by the

event horizon[5], and to the past by the ingoing time y^- solving (4.39). If we throw in a message which falls behind the horizon before this time its information is lost to the system, and obversely, is retrievable from the Hawking radiation in the bath, combined with the purifying system (the other side of the black hole). The time delay apparent in (4.39) is precisely the scrambling time, with $\beta = \frac{e^{\frac{k}{2}y^-}}{T_1}$ being the thermal scale at that time. Indeed, this delay may be written

$$t_{HP} = \frac{\beta}{2\pi} \log\left[\frac{16}{c}(S - S_0)\right], \tag{4.40}$$

with S the density of states at time u on the boundary. This result is precisely of the form (1.5) with coefficient $\alpha_{HP} = 1$ up to the interesting-but-tiny correction $\frac{\beta}{2\pi}\log\left(\frac{8}{c}\right)$. We find that the scrambling time is slightly reduced large c, because more bulk degrees of freedom encode proportionally more information in the Hawking radiation. One is tempted to associate this correction with a c-dependent minimal δE in (1.4), but this remains to be understood in detail.

We now compute the generalized entropy associated to these surfaces. Since the QES is very close to the event horizon, to leading order it suffices to evaluate the dilaton on the horizon $x^+ = t_\infty$ by the integral (2.31):

$$I(x^+ = t_\infty, x^-) = \int_0^{x^-} dt(x^- - t)(t_\infty - t)\{u, t\} \tag{4.41}$$

To do this, we use the same approach as before, using the fact that it approaches I_∞ as $x^- \to t_\infty$, and estimate the derivative at late times:

$$\begin{aligned}
\partial_- I(t_\infty, x^-) &= \int_0^{x^-} dt(t_\infty - t)\{u, t\} \\
&\sim \int^{x^-} \frac{dt}{2(t_\infty - t)} \\
&\sim -\tfrac{1}{2}\log(t_\infty - x^-) \\
\implies I(t_\infty, x^-) &\sim I_\infty + \tfrac{1}{2}(t_\infty - x^-)\log\left(\frac{t_\infty - x^-}{t_\infty}\right)
\end{aligned}$$

[5]This is the outside of the horizon in the event that evaporation continues to very late times. However, if we stop evaporation at some earlier time, the event horizon will be further out, due both to the necessary positive energy shock from turning off the coupling and the Hawking radiation that no longer is allowed to escape. Indeed, from (4.39) we see that the QES lags a scrambling time behind the boundary time so that an ingoing $O(1)$ positive energy pulse from switching off the coupling has a large effect on its location relative to the final future event horizon.

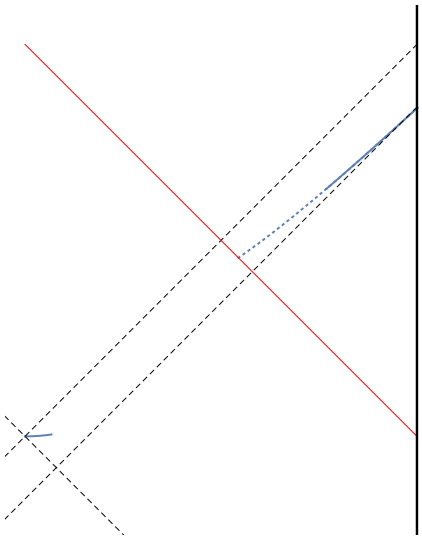

Figure 6. The location of quantum extremal surfaces as boundary time evolves (solid blue curves). The dashed lines indicate the original horizon before coupling to the bath, and the final event horizon. The red line indicates the shock at $x^- = 0$. The dotted blue line is the non-minimal quantum extremal surface before the Page time phase transition.

This gives us the dilaton value on the horizon, where the quantum extremal surface is located:

$$\phi \sim \bar{\phi}_r \left[2\pi T_1 + \tfrac{k}{2} \log \left(\frac{t_\infty - x^-}{t_\infty} \right) \right] \tag{4.42}$$

$$\sim \bar{\phi}_r 2\pi T_1 e^{-\frac{k}{2} y^-} \tag{4.43}$$

This gives the unsurprising result that the dilaton at the horizon matches the thermodynamic entropy associated with the energy at the corresponding ingoing time. As a further check, it is also consistent with the earlier result for the ∂_- variation of the dilaton. The bulk entropy gives a subleading contribution to S_{gen}; its gradient competes with the area term, but the quantity itself does not.

We note that an alternative way to perform the analysis at late times is to use the AdS isometries to change to a set of coordinates adapted to the time in question, which we discuss briefly in section 6.

The quantum extremal surface of section 4.2, which dominates at early times, gives a linearly increasing generalized entropy starting close to the entropy of the original black hole, with temperature T_0. The new surface we have found has genralized entropy which decreases linearly, but starting close to the thermodynamic entropy associated to the perturbed black hole after the shock, of temperature T_1. The latter is initially

larger, but at a time of order $t_P \sim k^{-1}\frac{T_1-T_0}{T_0}$ they exchange dominance, and our new surface becomes the quantum extremal surface. We interpret this phase transition with the Page time[6], when the black hole once again becomes old; see figure 6.

With these two families of quantum extremal surfaces, the generalized entropy perfectly reproduces the Page curve, indicative of unitary evaporation. The entropy increases at early times, since the outgoing Hawking radiation is entangled with the remainder of the system. This continues until the entropy of the black hole is close to its maximum at the given energy, at which point the entropy can only reduce, implying that the radiation is purified by early radiation (here including the left side black hole).

It is striking, and perhaps surprising, that we have recovered the Page curve from a semiclassical calculation. However, we emphasize that this does not resolve the information paradox, since it supposes that the generalized entropy of the quantum extremal surface correctly captures the fine-grained entropy of the black hole, without explaining the required correlations between early and late Hawking radiation. Indeed, the state of the matter in the bath at late times does not contain such correlations, and calculating the entropy of the bath from (3.26) points towards this loss of unitarity:

$$S_{\text{bath}} \sim \frac{\bar{\phi}}{4G_N} 4\pi T_1 (1 - e^{-\frac{k}{2}u}) \tag{4.44}$$

This increases for all time, at the rate given by the current temperature, approaching the entropy of a black hole at temperature $2T_1$.

Due to the large ground state entropy of R, there is no immediate contradiction between (3.26) and unitarity. In particular, the Araki-Lieb inequality is always satisfied. However, S_{bath} can be made arbitrarily large by successive iterations of injecting additional pulses of low-entropy energy into the black hole and further evaporation into the bath. In this way one demonstrate information loss as a violation of the Araki-Lieb inequality while remaining in the near-extremal regime throughout the evaporation. We will discuss such issues further in section 6.

4.6 Summary

Since the final discussion in section 6 will focus on more conceptual issues, we now pause to summarize the above technical results. In doing so, it is useful to take the viewpoint that begins with 3 systems L, R, B which at first do not interact. Here L, R are both AdS_2 black holes, and the black holes are highly entangled. The bath B begins in its ground state. This initial state is then perturbed by turning on a coupling between R and B, localized at the boundary of both systems. The coupling is then left on and becomes time-independent after a short initial transient.

[6] A.A. thanks G. Penington for discussion of this point.

The primary object of our study was the location of quantum extremal surfaces defined by the bulk entropy between an interior point and a point on the cut-off boundary at physical time u. Adding this entropy to $\frac{\phi_0+\phi}{4G}$ gives the generalized entropy, and for each u any QES are identified by extremizing the result with respect to the internal point. Our results then concern the evolution of such QES with u.

Before the coupling is turned on, there is a unique QES that (due to symmetry) agrees precisely with the classical extremal surface at the bifurcation surface of the original black hole horizon. After the coupling is turned on, this QES moves out toward the boundary in a spacelike direction. Several transient effects then occur whose effects on the entropy of the system were shown in figure 5. These effects are easily understood in terms of bulk quantum field theory with perturbative gravitational back-reaction. Switching on the coupling causes an initial sharp increase in entropy due to the production of short-distance entanglement between R and L. This is then followed by a decrease in entropy due to the escape of thermal radiation from R into the bath B. However, after a scrambling time $\frac{\beta_1}{2\pi}\log\left(\frac{E(T_1)}{E_S}\right)$ (4.12) one finds that the effective horizon of the black hole (say, the apparent horizon) has moved even further outward. As a result, the entropy defined by this QES begins to increase again as Hawking particles are emitted into the bath and their partners traveling inward add entropy to R. The resulting linear increase in entropy corresponds to the rising early entropy in the familiar Page curve of [45].

The entropy associated with this QES in fact continues to rise for all time. The corresponding entropy thus tracks precisely what one would expect if information were lost in evaporating black holes. However, since the actual entropy is described by the QES with the smallest S_{gen}, this same increase means that the dominant QES is likely to experience a phase transition at late times if we can identify another QES with smaller time derivative of S_{gen}.

Such a new QES can indeed be found emerging from the positive-energy shock that was produced at the AdS_2 boundary by turning on the coupling between R and B and which then falls into the black hole. The shock heats the black hole to a higher temperature T_1, associated with a higher energy E_1 and a higher Bekenstein-Hawking entropy S_1. The new QES is not a small perturbation to a classical extremal surface, but instead arises because there are surfaces in the classical geometry with sufficiently small expansions that they can be turned into a QES by quantum effects. Since it sits close to the new apparent horizon associated with the increased energy of the black hole, its entropy begins close to S_1, which at first significantly exceeds that of the first QES described above. This means that at the early times shown in figure (4.12), the new QES has larger entropy and thus is not relevant to determining the entanglement

wedge.

However, as the back hole continues to radiate and lose energy to B, the new QES remains close to the apparent horizon and thus slowly decreases in entropy. In particular, the actual value of S_{gen} at the new QES remains very close to that of the classical term $\frac{\phi_0+\phi}{4G}$, and thus close to the Bekenstein-Hawking density of states associated with the (time-dependent and now decreasing) energy of R. In other words, the S_{gen} of this QES agrees well with the entropy one would expect if one began with a maximally-entangled black hole and then watched it slowly evaporate via a unitary process. It is thus clear that at the relevant notion of the Page time the S_{gen} of this new surface will become less that of the other QES described above (whose S_{gen} continues to increase). This gives the stated phase transition of the dominant QES and then reproduces the decreasing part of the Page curve of [45].

It is also useful to note that the 2nd QES moves outward in a spacelike direction as time passes, though its motion becomes asymptotically null as the black hole reaches its final extremal equilibrium. The motion of both QES was shown above in figure 6. The worldline of the novel QES is marked as dotted in the region where it fails to dominate, and then solid where it dominates after the phase transition. This figure also shows the novel QES approaching the final event horizon from the interior at late times.

What is not shown in figure 6 is the relation between the physical boundary time u and the location of the corresponding QES on the solid blue curve. However, as discussed around equation (4.40) this relation becomes quite simple at late times. In that limit, the dominant QES for boundary time u lies on the ingoing null geodesic that was emitted from the boundary at a time $u - t_{HP}$ with

$$t_{HP} = \frac{\beta}{2\pi} \log \left[\frac{16}{c} (S - S_0) \right]. \tag{4.45}$$

This is a holographic manifestation of the Hayden-Preskill effect, where an ingoing signal disappears from the entanglement wedge of R after the time (4.45) and enters the entanglement wedge that would be defined by merging L and R; i.e., the signal can be recovered from the part of the system that describes the complement of the black hole into which the signal has fallen.

5 Left and Right Quantum Extremal Surfaces

For bulk matter in a pure state, a QES of a boundary region is identical to the QES of its complement. In particular, in a two-boundary geometry, the QES of the complete

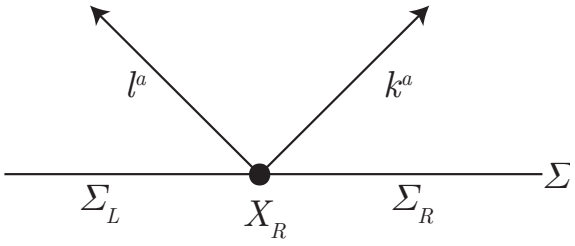

Figure 7. A Cauchy slice Σ containing the right quantum extremal surface X_R. By assumption, X_R splits Σ into two sides, labeled Σ_R and Σ_L. Σ_R is a Cauchy slice of the entanglement wedge of the right side, while Σ_R is not necessarily a Cauchy slice of the left entanglement wedge. The null vectors ℓ^a and k^a are normal to X_R. While this cartoon is two-dimensional, this setup is not restricted to any particular number of dimensions.

left boundary is identical to the QES of the complete right boundary whenever the full bulk state is pure.

In our situation, where the two-sided black hole in question has emitted some radiation into the bath, however, the bulk state is mixed: the right and left QESs are no longer required to coincide. Put differently, complementary recovery is not guaranteed when the bulk state is not pure [46]. This discrepancy between left and right QESs, which in turn corresponds to the gap between the left and right entanglement wedges, is an important aspect of our holographic realization of the Hayden-Preskill protocol.

Let us therefore investigate the left quantum expansion of the right QES in a general setting. We consider a generic bulk state in a two-sided (or in principle multi-boundary) geometry. We will temporarily work in general bulk dimension $d \geqslant 2$ and with bulk matter described by any local (not necessarily conformal) quantum field theory. Assuming, as we have thus far in this paper, that the right (minimal entropy) QES X_R lies in a globally hyperbolic region [7] which is well-described by an approximate geometry, we consider a Cauchy slice Σ of the bulk spacetime containing X_R. The right QES X_R splits Σ into two components: Σ_L, the left of X_R on Σ, and Σ_R, the right of X_R on Σ; see Fig. 7. This allows us to define the right and left reduced density matrices for bulk quantum fields: $\rho_{\Sigma_R} = \mathrm{tr}_{\Sigma_L}\rho$ and $\rho_{\Sigma_L} = \mathrm{tr}_{\Sigma_R}\rho$, which in turn define the right- and left- bulk von Neumann entropies of X_R: $S_R[X_R]$ and $S_L[X_R]$.

Because X_R is a right QES, it is by definition a stationary point of the generalized entropy functional: the variation of $S_R[X_R]$ due to shape deformations of X_R vanishes to first order in the deformation. Equivalently, the right quantum expansion of X_R in

[7] Here we mean the appropriate generalization to global hyperbolicity for asymptotically AdS spacetimes [47].

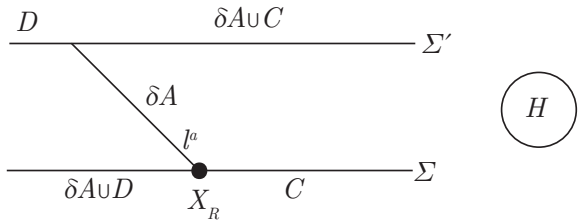

Figure 8. Here $\Sigma_R = C$ and $\Sigma_L = \delta A \cup D$. The infinitesimal deformation along ℓ^a is labeled δA and is parametrized by a choice of affine parameter λ_ℓ along the ℓ^a congruence. The ℓ^a congruence defines a similar split of the future Cauchy slice Σ' into the right side, $\delta A \cup C$ and the left side D.

any null direction vanishes. If ℓ^a and k^a are linearly independent null normals to X_R (in our $d = 2$ setup above, these may be chosen to be e.g. along $x_\pm$) with affine parameters λ_ℓ and λ_k respectively, we may schematically write this in the following way:

$$\frac{\delta S^{(R)}_{\text{gen}}[X_R]}{\delta \ell^a} = 0 \tag{5.1a}$$

$$\frac{\delta S^{(R)}_{\text{gen}}[X_R]}{\delta k^a} = 0, \tag{5.1b}$$

where $S^{(R)}_{\text{gen}}[X_R]$ is the right generalized entropy of X_R. Thus the right generalized entropy of X_R does not change to first order when X_R is infinitesimally deformed in a null direction (or, as noted above, any direction; it will simply turn out to be useful to work with null vectors).

To compare the left- and right- quantum expansions of X_R (and thus ascertain whether X_R is a left QES, and if not, by how much it fails to be one), consider deforming X_R along the k^a or ℓ^a directions. Since the change in area or in the dilaton for $d = 2$ is independent of whether we evaluate it on the left or right side, the difference between the left and right quantum expansions is entirely due to the evolution of $S_R[X_R]$ and $S_L[X_R]$ along the null directions.

To track this evolution, we evolve to a new Cauchy slice Σ' to the future of Σ, and divide Σ' into left and right components using ℓ^a or k^a This is illustrated in Fig. 8 for ℓ^a.

Using the decomposition labeled in Fig. 8 in the ℓ^a direction as well as the fact that X_R is a *right* quantum extremal, we know that as we take δA to zero, parametrized by the affine parameter λ_ℓ along ℓ^a, we get:

$$\delta S_R[X_R] = S[C] - S[C \cup \delta A] = \mathcal{O}(\lambda^2). \tag{5.2}$$

To relate the change along λ_ℓ of S_L with that of S_R, we would like to compare $S[D \cup \delta A] - S[D]$ with $S[C] - S[C \cup \delta A]$. This can be done with a judicious application of strong subadditivity (SSA) [48] to the system consisting of $\{\delta A, C, H\}$, where H is the emitted radiation:

$$S[C \cup H] + S[C \cup \delta A] - S[\delta A \cup C \cup H] - S[C] \geqslant 0. \tag{5.3}$$

Using the fact that $\rho_{\delta A \cup C \cup H \cup D}$ is pure, this immediately yields the desired comparison:

$$S[D \cup \delta A] - S[D] \geqslant S[C] - S[C \cup \delta A] = \mathcal{O}(\lambda^2). \tag{5.4}$$

So we find that

$$\frac{\delta S^{(L)}_{\text{gen}}[X_R]}{\delta \ell^a} \geqslant \frac{\delta S^{(R)}_{\text{gen}}[X_R]}{\delta \ell^a} = 0, \tag{5.5}$$

with the opposite inequality for deformations along the k^a direction:

$$\frac{\delta S^{(L)}_{\text{gen}}[X_R]}{\delta k^a} \leqslant \frac{\delta S^{(R)}_{\text{gen}}[X_R]}{\delta k^a} = 0. \tag{5.6}$$

That is, X_R is, from the perspective of the left side, a quantum "untrapped" (also called "normal") surface unless the SSA (5.3) is saturated.

When is SSA saturated? It is saturated strictly when $\rho_{\delta A \cup C \cup H}$ is a so-called Markov state: then the full state can be recovered completely from either of its marginals $\rho_{\delta A \cup C}$ or $\rho_{C \cup H}$. In particular, there exists a (state-dependent) *recovery map* $\mathbb{I}_A \otimes \mathcal{R}_C$, where $\mathbb{I}_A$ is the identity on A and $\mathcal{R}_C$ is a map from states on C to states on $C \cup H$, such that [49]:

$$\mathbb{I}_A \otimes \mathcal{R}_C[\rho_{\delta A \cup C}] = \rho_{\delta A \cup C \cup H}. \tag{5.7}$$

That is, the right QES is a left QES whenever the radiation is sufficiently entangled with the right entanglement wedge that we can fully reconstruct (in the sense of the recovery map above) the state of the radiation by just knowing the state of the entanglement wedge. This is precisely the case e.g. when we throw the radiation back into the left CFT so that the full bulk state is pure: the two QESs coincide; there is no gap between the left and right entanglement wedges.

It thus seems reasonable to anticipate that when the right QES is very close to also being a left QES – such as might be the case when the two surfaces in question are close to one another [8] – that $\rho_{\delta A \cup C \cup H}$ would be very close to a Markov state by some measure. That is, we may expect that the radiation is sufficiently entangled with the

[8] Note that approximate extremality does not guarantee proximity of extremal surfaces: the left quantum expansions of X_R being close to zero is not enough to guarantee that there is a neaby left quantum extremal surface.

right entanglement wedge that we may approximate its state using the state-dependent maps above.

In the context of type I von Neumann algebras, this can be made precise. The failure of saturation of strong subadditivity is bounded from below [50, 51]:

$$S[C\cup H]+S[C\cup\delta A]-S[\delta A\cup C\cup H]-S[C]\geqslant -2\log\left(\sup_{\mathcal{R}_C} F\left[\rho_{\delta A\cup C\cup H}|\mathbb{I}_A\otimes\mathcal{R}_C(\rho_{\delta A\cup C})\right]\right). \tag{5.8}$$

Here F is the fidelity between the state $\rho_{\delta A\cup C\cup H}$ and a state we obtain from a recovery map acting on $\rho_{\delta A\cup C}$, $\mathbb{I}_A\otimes\mathcal{R}_C(\rho_{\delta A\cup C})$; the supremum is taken over all recovery maps (this is sometimes termed "fidelity of recovery" [52, 53]). This suggests that if our right QES is close to being a left QES, then our state is very well approximated by a Markov state, the radiation H is entangled mostly with the right entanglement wedge, and there exists a state-dependent recovery map giving a good approximation of H from C. More precisely, $\rho_{\delta A\cup H}$ is close (as measured by the fidelity) to a product state.

Thus far our discussion in this section has been for general holographic systems. Let us now carry out the explicit calculation comparing the left and right quantum expansions of the right QES for our $D = 2$ black hole with conformal matter (we will see that SSA is very far from saturation in our system). For simplicity, we focus on the late (but not very late) time region where the (minimal entropy) right QES X_R lies entirely to the future of the shock (as illustrated in Fig. 9). To evaluate the left quantum expansions in the x^+ and y^- directions, consider a partial Cauchy slice extending from X_R to the left boundary. Unitarity of the bulk theory allows us to wiggle the Cauchy slice in any way we want, and for convenience we break it up into a null, constant x_+ component to the future of the shell and a component to the past of the shell. By Eq. 3.24, the renormalized entropy of intervals to the past of the shell in the Poincaré vacuum is independent of the interval, so it does not contribute to the derivative of $S_L[X_R]$ in either the x^+ or y^- directions.

We first consider the left- and right- quantum expansions in the y^- direction, which in the discussion above corresponds to the k^a direction. This calculation is particularly simple, since the the right interval can be broken into two null intervals: from $x_b = f(y_b)$ to (x^+, y^b) and from (x^+, y^b) to (x^+, y^-), the coordinates of X_R. The interval of interest in for the left quantum expansion, which we shall call D, extends from $(x^+, 0)$ to (x^+, y). This is illustrated in Fig. 9. Using (3.29), we immediately obtain the following expressions for the difference in the expansion along y^- is:

$$\partial_{y^-} S^{(R)}_{\text{gen}}[X_R] - \partial_{y^-} S^{(L)}_{\text{gen}}[X_R] = \frac{cy_b}{12y(y_- - y_b)} < 0, \tag{5.9}$$

where the inequality follows from the fact the by assumption the right QES lies to the future of the shock, and $0 < y^- < y_b$ (as a sanity check, this is consistent with the

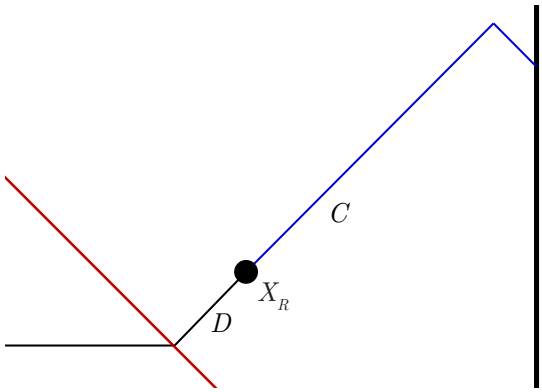

Figure 9. Breaking up the right interval and left intervals into null components for the calculation of the change in the entropy of X_R with variations along y_-. The red line is the shell.

sign expected from strong subadditivity). We note that as desired, the inequality is *not* saturated: the right QES fails to be a left QES (in fact, the difference is large). As explained above, we may interpret this as a manifestation of the fact that the radiation is far less entangled with the right half of the black hole than what would be necessary for us to recover even approximately the state of the radiation purely from knowing the state on this side of the black hole.

For completeness, we will also compute the difference in left- and right- quantum expansions in the x_+ direction, which in the above conventions corresponds to the ℓ^a direction. Using the same decomposition of the left interval as above, we obtain:

$$\partial_{x_+} S^{(R)}_{\text{gen}}[X_R] - \partial_{x_+} S^{(L)}_{\text{gen}}[X_R] = \frac{c}{12(x - x_b)} > 0. \tag{5.10}$$

Again the inequality (which has the correct sign dictated by strong subadditivity) fails to saturate by a large amount, as expected by the considerations above.

Finally, we have here used the fact that the failure of the right QES to achieve left quantum extremality is parametrized by the failure of a state to be Markovian; this is bounded from below by the failure of recovery maps on the right entanglement wedge to approximate the state of the radiation (where this failure is measured by the fidelity). It is tempting to use this non-saturation of strong subadditivity for generic states to attempt to prove that *in general* right and left QESs do not coincide. However, this is where we must be careful with the discreteness of the bound in (5.8). To prove that (5.3) is not saturated, we must show that the failure of SSA to saturate scales no faster with λ than $\mathcal{O}(\lambda)$ as we shrink δA to zero. If it goes to zero any faster, the first derivative – i.e. the quantum expansion – will not be sensitive to it. The bound (5.8) applies to discrete systems and thus contains no immediate information

on a continuously shrinking region. It is only indicative of the statement that the inequality is not, generically, saturated to *all* orders in λ. Even in the discrete model, it is unclear to us whether the nonsaturation will be linear in λ (though the fact that the sign of the inequality flips when $\lambda \to -\lambda$ does suggest that the a linear scaling is consistent). It may be possible to use different techniques to find the scaling order; we leave this to future work.

6 Discussion

The main outcome of our work was to show that Quantum Extremal Surfaces (QES) accurately described the expected unitary evaporation of black holes. While we focused on a particular model in low dimensions, we will shortly describe why a similar story must occur in much more general contexts. Crucially, we identified a novel QES that cannot be described as a small quantum correction to a classical extremal surface. Instead, this QES arises only due to quantum effects. We will return to the apparent tension in this statement below, but for the moment comment only that this new QES begins to dominate at the Page time, and that the corresponding quantum-HRT phase transition is directly responsible for the fact that our entropy decreases after the Page time as predicted by unitarity. See also section 4.6 for a succinct summary of our technical results.

We now reiterate these results with an eye toward interesting conceptual issues and future directions. The objective of sections 2-4 was to track the QES under one-sided evaporation of a two-sided thermofield-double-like black hole with initial temperature T_0. At some finite time we turned on a coupling that allowed the right side R of our system to freely radiate into a bath B. Although we worked only at the level of perturbative quantum corrections to classical solutions, the QES behaved in ways consistent with general expections from non-perturbative unitary evolution. In particular, the initial switching on of the coupling induces correlations between R and B which increase the entropy of R and also S_{gen} of the right QES by an amount S_S. The switching operation also injects energy E_S into our black hole, with $E_S \gg T_0$ so that the density of states of R now significantly exceeds $S_{\rm gen}$ of the QES. But this new energy takes some time to scramble, so as shown in figure 5 for a scrambling time (with coefficient $\alpha_S = 1$ the system proceeds as if this new phase space had not opened up. In particular, $S_{\rm gen}$ at the QES decreases for a time $\frac{\beta_0}{2\pi} \log \frac{E}{E_S}$ (4.12) as expected from (1.4) with coefficient $\alpha_S = 1$), and then increases again as the emission of Hawking radiation after this scrambling time creates more entanglement between R and B.

Further emission of Hawking radiation continues to cause the generalized entropy of the QES to increase while the energy of R and thus its density of states decreases.

Unitarity would require this behavior to change at what we may call the Page time for this setup when the two become equal. Thus far the QES itself has moved continuously in a spacelike but nearly-null direction[9]. But at this Page time there is a phase transition such that continuous deformations of the original QES no longer have minimal $S_{\rm gen}$.

Instead, another family of QESs becomes minimal after the Page time. The new family lies closer to the boundary than the ingoing pulse of energy E_S; see figure 6. In this new family of QESs, $S_{\rm gen}$ is a *decreasing* function of time that largely tracks the density of states of R as determined by its energy. Indeed, in the asymptotic future, $S_{\rm gen}$ becomes equal to the ground-state entropy of R up to a term associated with the coupling to B. Furthermore, this QES lies close to the future event horizon of R, and has an ingoing null coordinate that at late time lags that of the time-evolving boundary by precisely $t_{HP} = \frac{\beta}{2\pi} \log \left[\frac{16}{c}(S - S_0)\right]$ (4.40), where S is the density of states of R at the given time and S_0 is the ground state entropy. As described in section 1.1, such a lag is to be expected from the holographic description in [24] of the Hayden-Preskill protocol [23]. Note that the above scrambling time is not just as a rough timescale in our calculation, but as a sharp threshold. In particular, we find the coefficient to be $\alpha_{HP} = 1$ in (1.5). We may thus say that it gives the Hayden-Preskill delay time for our system in the limit of small messages.

It is tempting to associate the correction $\frac{\beta_0}{2\pi} \log \frac{16}{c}$ to (1.5) with a c-dependent minimum δE in (1.4), but this remains to be understood in detail. It would also be very interesting to reproduce both this term and the coefficients $\alpha_{HP} = 1$ and $\alpha_S = 1$ from a calculation in an SYK model [54–56] or the like. It would also be interesting to compare our results with analogous calculations for higher dimensional black holes. On general grounds one expects similar behavior, but one would like to understand the extent to which the coefficients $\alpha_{S\,\rm min} = \alpha_{HP} = 1$ and $\frac{16}{c}$ found here are universal and the extent to which they depend on properties of the black hole.

The fact that bulk entropy, which for us generally does not exceed $O(\log G_N)$, could significantly affect the location of the QES was due to the presence of large gradients in this entropy. Such large gradients arise naturally in the context of evaporating black holes from the well-known large boosts. It would be interesting to understand if such large gradients in entropy might somehow invalidate the semiclassical approximation used here, though we see no immediate reason for this to be the case[10]. Indeed, the size

[9]The fact that the QES moves in a spacelike or nearly-null direction towards the boundary R provides an important consistency condition for the case at hand where the extracted energy from R is immediately deposited in L as it precludes any bulk causal signaling from the left entanglement wedge to the right entanglement wedge.

[10]We thank D. Harlow for discussion on this point.

of such gradients is coordinate dependent, and with an appropriate choice of outgoing coordinate, the gradient of the entropy is of order one, and the gradient of the dilaton is suppressed by a factor of G_N.

This occurs naturally if we choose coordinates adapted to the boundary time of interest, in which case gradients of the dilaton a scrambling time in the past are suppressed by the familiar exponential divergence of trajectories near the horizon. One such set of coordinates is obtained by applying an AdS_2 isometry (an $SL(2,\mathbb{R})$ transformation γ, defining $\tilde{t} = \gamma(t)$, $\tilde{x}^{\pm} = \gamma(x^{\pm})$), a Rindler boost chosen such that $\tilde{t} = 0$ corresponds to a proper time u_0 of order k^{-1}. In the new $\tilde{x}^{\pm}$ coordinates[11], the metric retains the same form (3.1), and the dilaton profile is approximately given by the static black hole solution (2.11) with temperature $\tilde{T} = e^{-ku_0/2}T$, reflecting the fact that black hole is evolving adiabatically. This approximation to the dilaton profile is valid for a range of ingoing times of order k^{-1}, so in particular remains a good approximation a scrambling time ($\sim \log k^{-1}$) in the past, where the QES resides. In such coordinates, the QES lies near the would-be classical bifurcation surface at $\tilde{x}^+ = \frac{1}{\pi\tilde{T}}$, $\tilde{x}^- = -\frac{1}{\pi\tilde{T}}$ of the comparison static spacetime, though the actual evolving spacetime lacks a classical extremal surface in this region since the approximation for the dilaton breaks down for sufficiently small $\tilde{x}^- + \frac{1}{\pi\tilde{T}}$. This perspective makes it particularly clear that the existence of the QES is rather universal, being insensitive to the history of the black hole after a few scrambling times.

The transition of the QES after the Page time is directly related to the growing gap between the left and right QESs, which we expect is a consequence of the growing entropy of the bath. Indeed, this expectation can be sharpened in the situation where our system can be well-approximated by a type I von Neumann algebra, the amount by which the right QES fails to be a left QES is related to the failure of any (state-dependent) map acting purely on the right CFT to approximate the state of the radiation. This failure can be directly attributed to the entanglement of the bath with the left CFT: the inaccuracy of the approximation is a consequence of the density matrix ρ_{LB} not factorizing. One may speculate that the spacetime in between the two entanglement wedges may (in some appropriate sense) be emergent from entanglement with the bath.

Let us now discuss similar considerations in more general evaporating black holes. A model of this type closely related to our calculations above is given by starting with

[11] Explicitly, we can choose $\frac{du}{d\tilde{t}} = 1$ and $\frac{d^2u}{d\tilde{t}^2} = 0$ at the time $t = t_0$, $u = u_0$, in which case we find

$$\tilde{t} = \gamma(t) \sim \frac{1}{\pi\tilde{T}} \frac{t - t_0}{2t_\infty - t_0 - t}$$

our thermofield double state and and some time turning on couplings to a pair of auxiliary baths B_L, B_R. Both B_L, B_R begin in their ground state. The coupling to B_R is precisely as above and involves only our right system R, but we also introduce a corresponding coupling of B_L to R. Bulk causality then requires that the computation of the right quantum extremal surface is as before, and that the left quantum extremal surface behaves similarly. The joint system LR is a black hole that begins in a pure state (the thermofield double), radiates into B_L and B_R, and initially increases in entropy. In the semiclassical description of the bulk dual to LR we see this increase in entropy through the increase of entropy of bulk quantum fields. Indeed, the boundaries are homologous to the empty set, for which the generalized entropy at order $1/G_N$ is precisely the bulk entropy on a complete Cauchy slice through the bulk spacetime.

However, an interesting transition occurs if such radiation continues past the Page time, where the bulk entropy $S_{\rm bulk}$ on our Cauchy slice becomes greater than the density of states S_{LR} of LR. As this requires the original Bekenstein-Hawking entropy to exceed the ground state entropy by more than a factor of 2, remaining in the near-extremal limit where our JT model is a controlled approximation to known dualities requires that this be achieved by repeatedly injecting a large number of low-entropy pulses of energy into the bulk and then waiting for the bulk to Hawking radiate this into the bath before sending in the next pulse, though one may alternatively study higher-dimensional AdS models where controlled dualities describe far-from-extremal black holes. In either case, at late times we will find a new QES close to each horizon, since the relevant regions of the spacetime become adiabatically close stationary black holes so that we may again apply our earlier considerations.[12]. Since energetic considerations will force $\frac{A}{4G_N} \approx S_{LR}$, beyond the Page time one finds $S_{\rm bulk} > \frac{A}{4G_N}$, so the quantum extremal surface with minimal generalized entropy is then near the horizon and is no longer the empty surface. In this way the quantum extremal surface undergoes a first-order

[12]One may also give an alternative argument using a maximin definition of the quantum extremal surface analogous to the classical HRT surface discussion in [2]; the arguments of [2] also apply to 1+1 quantum extremal surfaces so long as one assumes the quantum focusing conjecture of [44] (though there are some subtleties in the choice of Cauchy slice used to define the entropy in the quantum focusing conjecture when the black hole is an open system). Suppose that one turns off the coupling at some late time, constructs a new spacetime from the appropriate late-time Cauchy data in the original spacetime, and studies $S_{\rm gen}$ for surfaces in the resulting new spacetime. The new spacetime is essentially AdS-Schwarzschild outside the horizon, but has a long one-sided wormhole (with a large causal shadow) inside. On any Cauchy surface, there will be *some* surface not too far inside the black hole and satisfying the homology constraint with area less than $4G_N S_{XL}$, and thus with generalized entropy also close to or smaller than $4G_N S_{XL}$. In contrast, as above the empty set is associated with greater generalized entropy and so cannot be the maximim surface. Instead, the maximin surface will must have generalized entropy bounded above by $4G_N S_{XL}$ up to logarithmic corrections.

phase transition at the Page time. Because the vast majority of the bulk entropy will be localized inside the black hole and local effects from quantum fields are small, the entropy of this second QES will be given by the horizon area $\frac{A}{4G_N}$ up to logarithmic corrections

Similar comments clearly apply to one-parameter families of so-called 'bag-of-gold' spacetimes with large regions of spacetime behind a black hole horizon in which the bulk quantum fields are placed in a mixed state of large entropy. For $S_{\rm bulk} < \frac{A}{4G_N}$, the minimal quantum extremal surface is the empty set, but for $S_{\rm bulk} > \frac{A}{4G_N}$ it jumps to near the black hole horizon; see [24] for recent examples of this phenomenon in the context of the SYK model. In this context, both quantum extremal surfaces are in fact near classical extremal surfaces. In both the small and large $S_{\rm bulk}$ regimes, the dominant quantum extremal surface indicates that the entanglement wedge is associated with an entropy less than or equal to the density of states of the system to which it is dual.

Returning to the evaporating black hole, we see that tracking the quantum extremal surface using only perturbative semiclassical dynamics in the bulk fully reproduces the expected Page curve, including in particular the decay to zero of the entropy if the system decays to a non-degenerate ground state. And it is also interesting that it does so by terminating the entanglement wedge within a short distance of the horizon of the remaining late-time black hole.

This observation will surely fan the flames of black hole information debates and on the possible role of firewalls [57–59] in particular. On one hand, the termination of the entanglement wedge at the edge of the black hole may indicate that no meaningful spacetime exists farther in. On the other, the fact that a purely perturbative semi-classical model of the bulk defines quantum extremal surfaces that reproduce the expected Page curve may indicate that the interior spacetime *is* meaningful, that no firewall is needed, and that the interior is somehow dual to the (arbitrary!) bath system (see e.g. [60, 61] for related ideas). Indeed, if the bath were holographic and described by a bulk that connected to this interior by even a very small wormhole, this would be the natural conclusion of studying the QES for the bath. While under normal circumstances the minimal QES would lie at the small wormhole (effectively the trivial surface discussed above), beyond the Page time the large bulk entropy in the bath system would move the QES out to the horizon. In this context, for each boundary time the setting would be much like that of the original ER=EPR discussion [61] with similar potential implications for black hole information.

However, such an interpretation would return us to the familiar problem that a perturbative semi-classical bulk description of the radiation does not provide the correlations between Hawking quanta necessary to purify the emitted radiation. In particular, in the model described above one can use perturbative semi-classical bulk physics

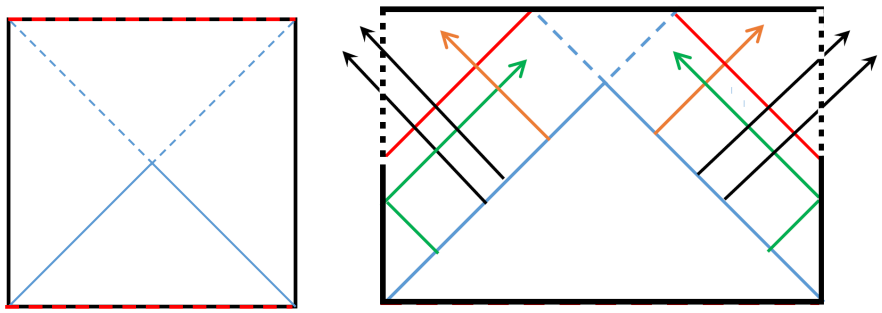

Figure 10. (Left) For an eternal black hole dual to the thermofield double state, the segments of the horizons to the past of the bifurcation surface (solid lines) join to form a mostly-null Cauchy surface. Data in this quantum state then propagates outward to the right and left. (Right) Turning on a coupling allows some modes from the above Cauchy surface to escape. But those that reach the boundary before the coupling is turned on are reflected back into the singularity, and modes close to the original bifurcation surface are focused by an incoming pulse of positive energy (red lines without arrows) associated with turning the coupling on. As a result of this focussing, semiclassical physics sates that such modes also fail to escape.

to track precisely the flow of entropy and information into $B_L B_R$ as was done for B_R in section 4. But since the modes on both right and left that escape into B_L, B_R remain entangled with both modes that are reflected back into the black hole before the coupling is turned on and those that lie to the future of all modes that escape, doing so would find $B_L B_R$ to end in a highly mixed state; see figure 10. This is directly analogous to the Araki-Leib violation found for our one-sided evaporation at the end of section 4 in the context where one executes repeated cycles of exciting the near-extremal JT-bulk by a small amount and then letting it Hawking radiate into the bath.

In both cases, unitarity thus requires that bulk semi-classical physics fails to correctly compute the late-time entropy of the bath. A similar failure to correctly compute the entropy of bulk quantum fields themselves is suggested by noting that in the full bulk spacetime the predicted entropy of such fields exceeds the density of states in the dual CFT. Indeed, since this is precisely the feature that led to the QES phase transition at the Page time, it is interesting to ask if such a phase transition really occurs in a full non-perturbative treatment. A plausible alternative speculation might be that the non-perturbative system instead evolves so as to become extremely close to this phase transition at late times – perhaps even close enough that the concept of a definite entanglement wedge ceases to be well-defined. On a positive note, however, since the entropy of semi-classical bulk fields within the entanglement wedge associated with the

QES after the Page time appears consistent with the dual CFT density of states, it is at least self-consistent to use semi-classical physics within this wedge and to suppose that non-perturbative corrections become large only when one probes more deeply into the bulk.

We thus find that any perspective continues to lead to many open questions. In order to make real progress in such debates it seems critical to understand more precisely what is meant by duality between field theory degrees of freedom and an entanglement wedge in the bulk. Thinking of the entanglement wedge as defining the bulk region that can be reconstructed from the stated degrees of freedom suggests this be done by further investigating the role of quantum error correction and recovery maps in gravitational holography. We thus look forward to further progress on this front, or on other related aspects of holographic duality.

Acknowledgements We wish to thank D. Harlow, A. Harrow, Z. Komargodski, N. Lashkari, J. Maldacena, and D. Stanford. AA is supported by funds from the Ministry of Presidential Affairs, UAE. N.E. is supported by the Princeton University Gravity Initiative and by NSF grant No. PHY-1620059. D.M. acknowledges support from NSF grant PHY1801805 and funds from the University of California. Support for H.M. was provided by a DeBenedictis Postdoctoral Fellowship and by funds from the University of California. H.M. would like to thank C. sinensis for moral support.

References

[1] 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].

[2] A. C. Wall, *Maximin Surfaces, and the Strong Subadditivity of the Covariant Holographic Entanglement Entropy*, *Class. Quant. Grav.* **31** (2014), no. 22 225007, [arXiv:1211.3494].

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

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

[5] M. Headrick and T. Takayanagi, *A Holographic proof of the strong subadditivity of entanglement entropy*, *Phys. Rev.* **D76** (2007) 106013, [arXiv:0704.3719].

[6] A. Almheiri, X. Dong, and D. Harlow, *Bulk Locality and Quantum Error Correction in AdS/CFT*, *JHEP* **04** (2015) 163, [arXiv:1411.7041].

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

[8] 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), no. 2 021601, [arXiv:1601.05416].

[9] T. Faulkner and A. Lewkowycz, *Bulk locality from modular flow*, *JHEP* **07** (2017) 151, [arXiv:1704.05464].

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

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

[12] A. Lewkowycz and J. Maldacena, *Generalized gravitational entropy*, *JHEP* **08** (2013) 090, [arXiv:1304.4926].

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

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

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

[16] R. M. Wald, *Black hole entropy is the Noether charge*, *Phys. Rev.* **D48** (1993), no. 8 R3427–R3431, [gr-qc/9307038].

[17] V. Iyer and R. M. Wald, *A Comparison of Noether charge and Euclidean methods for computing the entropy of stationary black holes*, *Phys. Rev.* **D52** (1995) 4430–4439, [gr-qc/9503052].

[18] T. Jacobson, G. Kang, and R. C. Myers, *On black hole entropy*, *Phys. Rev.* **D49** (1994) 6587–6598, [gr-qc/9312023].

[19] X. Dong, *Holographic Entanglement Entropy for General Higher Derivative Gravity*, *JHEP* **01** (2014) 044, [arXiv:1310.5713].

[20] R.-X. Miao and W.-z. Guo, *Holographic Entanglement Entropy for the Most General Higher Derivative Gravity*, *JHEP* **08** (2015) 031, [arXiv:1411.5579].

[21] W. H. Zurek, *Entropy Evaporated by a Black Hole*, *Phys. Rev. Lett.* **49** (1982) 1683–1686.

[22] D. N. Page, *COMMENT ON 'ENTROPY EVAPORATED BY A BLACK HOLE'*, *Phys. Rev. Lett.* **50** (1983) 1013.

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

[24] A. Almheiri, *Holographic Quantum Error Correction and the Projected Black Hole Interior*, arXiv:1810.02055.

[25] D. Harlow, *Jerusalem Lectures on Black Holes and Quantum Information*, *Rev. Mod. Phys.* **88** (2016) 015002, [arXiv:1409.1231].

[26] D. Marolf, *The Black Hole information problem: past, present, and future*, *Rept. Prog. Phys.* **80** (2017), no. 9 092001, [arXiv:1703.02143].

[27] G. Penington, *Entanglement Wedge Reconstruction and the Information Paradox*, arXiv:1905.08255.

[28] S. Leichenauer, *Disrupting Entanglement of Black Holes*, *Phys. Rev.* **D90** (2014), no. 4 046009, [arXiv:1405.7365].

[29] P. Gao, D. L. Jafferis, and A. Wall, *Traversable Wormholes via a Double Trace Deformation*, *JHEP* **12** (2017) 151, [arXiv:1608.05687].

[30] J. Maldacena, D. Stanford, and Z. Yang, *Diving into traversable wormholes*, *Fortsch. Phys.* **65** (2017), no. 5 1700034, [arXiv:1704.05333].

[31] J. Maldacena, D. Stanford, and Z. Yang, *Conformal symmetry and its breaking in two dimensional Nearly Anti-de-Sitter space*, *PTEP* **2016** (2016), no. 12 12C104, [arXiv:1606.01857].

[32] J. Engelsöy, T. G. Mertens, and H. Verlinde, *An investigation of AdS_2 backreaction and holography*, *JHEP* **07** (2016) 139, [arXiv:1606.03438].

[33] A. Almheiri and B. Kang, *Conformal Symmetry Breaking and Thermodynamics of Near-Extremal Black Holes*, *JHEP* **10** (2016) 052, [arXiv:1606.04108].

[34] S. Sachdev, *Universal low temperature theory of charged black holes with AdS_2 horizons*, 2019. arXiv:1902.04078.

[35] A. Almheiri and J. Polchinski, *Models of AdS_2 backreaction and holography*, *JHEP* **11** (2015) 014, [arXiv:1402.6334].

[36] K. Jensen, *Chaos in AdS_2 Holography*, *Phys. Rev. Lett.* **117** (2016), no. 11 111601, [arXiv:1605.06098].

[37] J. Maldacena and X.-L. Qi, *Eternal traversable wormhole*, arXiv:1804.00491.

[38] G. J. Galloway and M. Graf, *Rigidity of asymptotically $AdS_2 \times S^2$ spacetimes*, arXiv:1803.10529.

[39] T. Faulkner, R. G. Leigh, O. Parrikar, and H. Wang, *Modular Hamiltonians for Deformed Half-Spaces and the Averaged Null Energy Condition*, *JHEP* **09** (2016) 038, [arXiv:1605.08072].

[40] P. Calabrese and J. Cardy, *Entanglement and correlation functions following a local quench: a conformal field theory approach*, *J. Stat. Mech.* **0710** (2007), no. 10 P10004, [arXiv:0708.3750].

[41] P. Calabrese and J. Cardy, *Entanglement entropy and conformal field theory*, *J. Phys.* **A42** (2009) 504005, [arXiv:0905.4013].

[42] C. T. Asplund and A. Bernamonti, *Mutual information after a local quench in conformal field theory*, *Phys. Rev.* **D89** (2014), no. 6 066015, [arXiv:1311.4173].

[43] I. Affleck and A. W. W. Ludwig, *The fermi edge singularity and boundary condition changing operators*, *Journal of Physics A: Mathematical and General* **27** (aug, 1994) 5375–5392.

[44] R. Bousso, Z. Fisher, S. Leichenauer, and A. C. Wall, *Quantum focusing conjecture*, *Phys. Rev.* **D93** (2016), no. 6 064044, [arXiv:1506.02669].

[45] D. N. Page, *Average entropy of a subsystem*, *Phys. Rev. Lett.* **71** (1993) 1291–1294, [gr-qc/9305007].

[46] D. Harlow, *The Ryu?Takayanagi Formula from Quantum Error Correction*, *Commun. Math. Phys.* **354** (2017), no. 3 865–912, [arXiv:1607.03901].

[47] R. P. Geroch, *The domain of dependence*, *J. Math. Phys.* **11** (1970) 437–439.

[48] E. Lieb and M. Ruskai, *Proof of the strong subadditivity of quantum-mechanical entropy*, *J.Math.Phys.* **14** (1973) 1938–1941.

[49] P. Hayden, R. Jozsa, D. Petz, and A. Winter, *Structure of States Which Satisfy Strong Subadditivity of Quantum Entropy with Equality*, *Communications in Mathematical Physics* **246** (Jan, 2004) 359–374, [quant-ph/0304007].

[50] O. Fawzi and R. Renner, *Quantum Conditional Mutual Information and Approximate Markov Chains*, *Communications in Mathematical Physics* **340** (Dec, 2015) 575–611, [arXiv:1410.0664].

[51] D. Sutter, O. Fawzi, and R. Renner, *Universal recovery map for approximate Markov chains*, *Proceedings of the Royal Society of London Series A* **472** (Feb, 2016) [arXiv:1504.07251].

[52] M. Berta and M. Tomamichel, *The Fidelity of Recovery is Multiplicative*, arXiv:1502.07973.

[53] K. P. Seshadreesan and M. M. Wilde, *Fidelity of recovery, squashed entanglement, and measurement recoverability*, *Phys. Rev. A.* **92** (Oct, 2015) 042321, [arXiv:1410.1441].

[54] A. Kitaev, *A simple model of quantum holography- 2015. Talks at KITP, April 7,and May 27*, .

[55] J. Maldacena and D. Stanford, *Remarks on the Sachdev-Ye-Kitaev model*, *Phys. Rev.* **D94** (2016), no. 10 106002, [arXiv:1604.07818].

[56] 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].

[57] A. Almheiri, D. Marolf, J. Polchinski, and J. Sully, *Black Holes: Complementarity or Firewalls?*, *JHEP* **02** (2013) 062, [arXiv:1207.3123].

[58] A. Almheiri, D. Marolf, J. Polchinski, D. Stanford, and J. Sully, *An Apologia for Firewalls*, *JHEP* **09** (2013) 018, [arXiv:1304.6483].

[59] D. Marolf and J. Polchinski, *Gauge/Gravity Duality and the Black Hole Interior*, *Phys. Rev. Lett.* **111** (2013) 171301, [arXiv:1307.4706].

[60] E. Verlinde and H. Verlinde, *Behind the Horizon in AdS/CFT*, arXiv:1311.1137.

[61] J. Maldacena and L. Susskind, *Cool horizons for entangled black holes*, *Fortsch. Phys.* **61** (2013) 781–811, [arXiv:1306.0533].