PAPER • OPEN ACCESS

Hawking radiation correlations of evaporating black holes in JT gravity

View the article online for updates and enhancements.

Hawking radiation correlations of evaporating black holes in JT gravity

Timothy J Hollowood[1,*] , **S Prem Kumar**[1] **and Andrea Legramandi**[2]

[1] Department of Physics, Swansea University, Swansea, SA2 8PP, United Kingdom
[2] Dipartimento di Fisica, Università di Milano-Bicocca, Piazza della Scienza 3, I-20126 Milano, Italy

E-mail: t.hollowood@swansea.ac.uk, s.p.kumar@swansea.ac.uk and a.legramandi@campus.unimib.it

CrossMark

Abstract

We consider the Hawking radiation emitted by an evaporating black hole in JT gravity and compute the entropy of arbitrary subsets of the radiation in the slow evaporation limit, and find a zoo of possible island saddles. The Hawking radiation is shown to have long range correlations. We compute the mutual information between early and late modes and bound from below their squashed entanglement. A small subset of late modes are shown to be correlated with modes in a suitably large subset of the radiation previously emitted as well as later modes. We show how there is a breakdown of the semi-classical approximation in the form of a violation of the Araki–Lieb triangle entropy inequality, if the interior of the black hole and the radiation are considered to be separate systems. Finally, we consider how much of the radiation must be collected, and how early, to recover information thrown into the black hole as it evaporates.

Keywords: black holes, Hawking radiation, JT gravity

1. Introduction

Recent work [1, 2] has led to a step change in understanding the information loss paradox of black holes. We can now see the missing ingredient in Hawking's calculation [3] using only semi-classical methods: there are new saddle points of the functional integral used for calculating entropies of quantum fields on the black hole geometry, the replica wormholes.

---

*Author to whom any correspondence should be addressed.

These new saddles allow one to calculate semi-classically the Page curve [4] of a black hole. A pure state of matter that collapses to form a black hole will return to a pure state. The replica wormholes, along with Page's original insight, imply that Hawking radiation must have very non-trivial correlations between subsets emitted at different times. It is the purpose of this paper to probe these correlations for an evaporating black hole in the Jackiw-Teitleboim (JT) gravity set up for which the replica wormhole calculations have fully developed.

Quantum information tools have provided a very powerful way to understand the behaviour of black holes. This goes back to tracking the entropy of the Hawking radiation across the lifetime of an evaporating black hole, the Page curve [4], or the way that the information of systems, e.g. a diary, dropped into a black hole, can be recovered in the radiation, considered by Hayden and Preskill [5]. The Page curve has been derived in these, or related, scenarios [1, 2, 6–8] (see also [9–11]). It is a goal of this work, to show that replica wormhole techniques, and the effective rules that they give rise to, mean that more refined information processing properties of black holes can also be calculated from first principles via standard quantum field theory calculations. For instance, we are able to derive the detailed aspects of information recovery anticipated by Hayden and Preskill. We will show that the information contained in a diary, in this case taking the form of a local quench in the Quantum Field Theory (QFT), thrown into a black hole before the Page time, can be recovered at a time

$$t = t_{\text{Page}} + \frac{1}{k} \cdot \frac{S_{\text{diary}}}{S_{\text{BH}}}, \qquad t_{\text{Page}} = \frac{2}{k} \log \frac{3}{2}, \tag{1.1}$$

where S_{diary} is the entropy of the diary assumed to be small compared with S_{BH}, the Bekenstein–Hawking entropy of the black hole, and k is the evaporation rate.

It has been appreciated for a while, that if black holes obey the laws of quantum mechanics then something quite dramatic must happen to reconcile unitarity of evaporation with the rules of effective theories. More precisely, if a Hawking mode B is emitted by an old black hole, one past the Page time, then it must be entangled with a mode of the early Hawking radiation R_B in order to ensure unitarity. On the other hand, the usual rules of local effective theory imply that the Hawking mode must be entangled with its partner mode behind the horizon A. Quantum mechanics, of course, forbids B to be maximally entangled with two separate subsystems. One reaction is to give up the entanglement across the horizon leading to a separable quantum state with a diverging energy density at the horizon, a 'firewall' [12]. Another, arguably even more dramatic answer that maintains a smooth geometry at the horizon, is to hypothesize that A and R_B are not separate subsystems $A = R_B$: modes on the inside of the black hole are actually living in the Hilbert space of the early Hawking radiation (see [13]).

One of the goals of this work is to pin down where R_B lies within the early radiation by taking B to be a small subset of modes emitted at a certain time by an old black hole. We then attempt to locate R_B by maximizing B's mutual information with the radiation emitted earlier[3]. We find that R_B must lie in a large subset of modes emitted from around the Page time to just before B. This means that the purifier R_B is de-localized in the earlier radiation. This is in tune with ideas from quantum information theory that suggest that extracting R_B would be computationally a hard problem [14].

The astonishing '$A = R_B$' scenario grew out of ideas of *black hole complementarity* [14–23] and ER = EPR [20] (see the review [13] for a detailed discussion and other references). If $A = R_B$ is really true, then it is legitimate to ask what goes wrong with conventional effective

---

[3] This is a necessary condition because the mutual information by itself does not imply entanglement. However, we also consider a genuine measure of entanglement known as the *squashed entanglement*. Half the mutual information is an upper bound for the latter.

quantum field theory on the black hole background? It is a goal of this work to show that one way that the breakdown of QFT manifests as a breakdown of the consistency conditions on the entropies of spatially separated regions of the quantum fields, specifically the triangle inequality of Araki and Lieb [24]

$$S_{AR} \geqslant |S_A - S_R|. \tag{1.2}$$

Here, A will be modes behind the horizon and R the Hawking modes of an old black hole. The reason for the breakdown will be that for an old black hole, S_R is dominated by a replica wormhole saddle (has an 'island') whereas S_A and S_{AR} are not. It is the island then that disrupts the usual consistency of the entropies in QFT. In retrospect, the conclusion is not surprising because the triangle inequality does not apply to subregions that overlap, and when R has an island then this overlaps with A.

   The organization of this paper is as follows. In section 2, we review the set up in JT gravity, where an evaporating black hole is created by a local quench [11] (related to scenarios in [6, 25]). In this section, we emphasize the simplifications that occur in the slow evaporation limit. In section 3, we describe how to evaluate the entropy of a set of intervals in the bath, including the island saddles that follow from the replica wormholes. The key computation is the solution for the island saddles that we show simplifies in the slow evaporation limit. There are a whole zoo of island saddles, even with the assumptions we make, but usually only a few are actually needed. The remaining sections put our entropy formulae to use. In section 4, we derive the Page curve of the evaporating black hole and extract the Page time. We then calculate the correlation between the early and late Hawking radiation in the form of the mutual information. This shows that there are strong correlations, as expected on the basis of Page's analysis. The correlations can be shown to be quantum, i.e. entanglement, by establishing a lower bound on the *squashed entanglement*, a measure of entanglement in mixed states. In section 5 we analyse the correlation between the early and late Hawking modes in more detail. We pick a narrow interval of late modes B and find out which interval of early modes it is maximally correlated with. This establishes that modes entangled with B, R_B, are de-localized over a large subset of the early modes that extends from around the Page time to the modes emitted just before B. This is what is expected: the entangled modes R_B should be difficult to extract from the early radiation [14]. Section 6 is devoted to showing that, when the interior of the black hole is considered, there is breakdown of the Araki–Lieb triangle inequality for the entropies of the interior and the radiation. This provides a smoking gun for the $A = R_B$ scenario. In section 7 we consider how information thrown into the black hole in the form of an entropy carrying local quench in the Conformal Field Theory (CFT) is recovered in the Hawking radiation. We find detailed agreement with the quantum information analysis of Hayden and Preskill [5]. Finally in section 8 we show an operator insertion behind the horizon is observable in the bath, if the appropriate interval in the bath is in its island saddle.

1.1. Entropy as an observable

For the new developments involving black holes and the information loss paradox, the entropy plays a key rôle. Our results involve the entropies of sub-regions in the radiation bath and associated quantum information measures. It is a natural question to ask whether these entropies are observable, even in principle? This is important because we have shown that the entropy of sub-regions of the bath are sensitive to physics behind the horizon and so it is fundamental to understand if it is actually observable from the bath.

   More generally, we can consider the Rényi entropies $S_A^{(n)} = (1 - n)^{-1} \log \text{tr} \, \rho_A^n$ of the sub-region, where $S_A = \lim_{n \to 1} S_A^{(n)}$. Note that for a finite dimensional subsystem A of dimension

d_A, only the first $d_A - 1$ Rényi entropies are needed to extract the eigenvalues of ρ_A and hence the von Neumann entropy of A.

The Rényi entropies are not directly conventional observables in the sense of being associated to a Hermitian operator. They can be computed by joint measurements on n copies of the system. However, at least in the case of finite dimensional systems, they can be measured by measuring a set of conventional, but random, observables on a single copy of the system (e.g. [26, 27]). The idea is take a complete set of rank-1 projection operators on A, Π_j, $j = 1, 2, \ldots, d_A$. Then define the rotated sets $U\Pi_j U^\dagger$, for an arbitrary unitary operator U on A. These are associated to some Hermitian operators $\mathcal{O}_U = \sum_j \lambda_j U\Pi_j U^\dagger$. Then one measures $\mathcal{O}_U$ in the conventional sense on copies of the system in order to estimate the Born rule probabilities for the outcomes j and for an arbitrary U,

$$p_U(j) = \text{tr}\left(\rho_A U\Pi_j U^\dagger\right). \tag{1.3}$$

The estimates for the Rényi entropies are then given as averages of powers of the probabilities in the unitary ensemble; for example, for the second Rényi entropy

$$\text{tr}\,\rho_A^2 = (d_A + 1)\sum_j \overline{p_U(j)^2} - 1, \tag{1.4}$$

where the over-line indicates an average over the unitary orientation. The higher $\text{tr}\,\rho_A^n$ involve a similar average of a polynomial of order n in the probability. The explicit formula is given in [27]:

$$\sum_j \overline{p_U(j)^n} = \sum C_{b_1,\ldots,b_n} \prod_{m=1}^{n} \left(\text{tr}\,\rho_A^m\right)^{b_m}, \tag{1.5}$$

where the sum is over conjugacy classes $1^{b_1} 2^{b_2} \cdots n^{b_n}$ of the symmetric group S_n and

$$C_{b_1,\ldots,b_n} = \frac{d_A n!}{\prod_{m=1}^{n} (d_A + m - 1) j^{b_m} b_m!}. \tag{1.6}$$

In a real application of this protocol, the average over the unitary ensemble is realized in terms of a discrete sampling known as a *k-unitary design*.

The conclusion is that, in principle, the entropy can be measured locally on a subsystem using conventional, albeit random, quantum measurements, at least for finite dimensional systems.
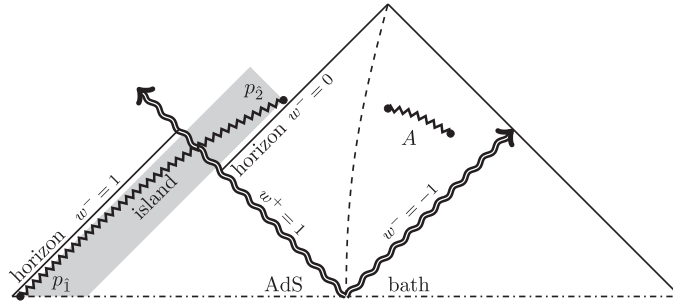
2. The evaporating black hole

The setup consists of the extremal black hole in Jackiw–Teitelboim gravity [28, 29] defined on a patch AdS$_2$, with the standard metric in Poincaré coordinates

$$ds^2 = -\frac{4\,dx^+\,dx^-}{(x^- - x^+)^2}, \tag{2.1}$$

with a half Minkowski space spliced on the time-like boundary to act as a radiation bath with metric $ds^2 = -dy^+ dy^-$ [2, 6, 8, 25, 30]. Then along the boundary $x^+ \sim x^-$,[4] we have $x^\pm = y^\pm$.

---

[4] The details of the regularization at the boundary are described in [25].

Figure 1. A Penrose diagram showing a shockwave inserted at the boundary of the AdS and bath regions with an in-going component that excites the extremal black hole, shifting the horizon out, leading to an evaporating black hole. Also shown is an interval A in the bath and its island, the shaded area in the AdS region, the causal domain of the two Quantum Extremal Surfaces (QES) QESs $p_{\hat{a}}$.

The boundary conditions are transparent, so that modes of the CFT propagate through the boundary without reflection. On the AdS region, the gravitational sector includes the dilaton, which in the extremal black hole case takes the form

$$\phi = \phi_0 + \frac{2\phi_r}{x^- - x^+}. \tag{2.2}$$

A CFT, which we take to be a large number of free fermions, propagates across the whole geometry, the AdS and Minkowski regions.

The evaporating black hole is created by a local quench [11]—an operator insertion—in the CFT initiated from a point on the boundary at $t = 0$ that leads to an in-going and out-going shockwave: see figure 1.[5] The quench corresponds to a CFT state created by the action of a local operator on the vacuum at the time-like boundary:

$$\mathcal{O}(y^\pm = i\varepsilon) |0\rangle, \tag{2.3}$$

where the small shift ε in the imaginary time direction is needed to ensure that the state can be normalized. The in-going component of the shockwave, $x^+ = 0$, carries energy into the black hole and excites it to a black hole of inverse temperature

$$\beta = \sqrt{\frac{\pi\phi_r}{4G_N E_{\text{shock}}}}, \tag{2.4}$$

where $E_{\text{shock}} = \Delta_{\mathcal{O}}/\varepsilon$. When ε is small, the shockwave energy is large and its energy–momentum tensor becomes concentrated on the two wavefronts $x^+ = 0$ and $y^- = 0$ propagating into, and away from, the black hole, respectively.

In the following, we will suppose that the intrinsic entropy of the shockwave between the in- and out-going components is vanishing, or at least small compared with the gravitational entropy, and can be ignored. For simplicity, we will also assume that the gravitational entropy

---

[5] We choose Penrose diagrams so that the straight line at the bottom corresponds to $t = 0$ in the bath and $t_{\text{Poincaré}} = 0$ in the AdS region. The boundary between the AdS region and the bath, $x^+ = x^-$, or $y^+ = y^-$, becomes curved behind the shockwaves.

of the excited black hole is much greater than the extremal entropy[6]

$$S_{\text{BH}} = \frac{\pi c}{6\beta k} \gg S_0 = \frac{\phi_0}{4G_N}. \tag{2.5}$$

The key physical quantity here is the inverse time scale

$$k = \frac{G_N c}{3\phi_r}, \tag{2.6}$$

which sets the rate of evaporation of the black hole. We shall work in the limit where $k \ll 1$ (relative to the AdS scale which is set to 1) with β fixed and time scales with kt fixed. This limit is simply one of expediency rather than necessity that means we can avoid having to resort to numerical techniques.

In front of the shockwave, the AdS and bath coordinates are related via $x^\pm = y^\pm$. We continue this through the shockwave by writing $x^\pm = f(y^\pm)$. The exact expression for $f(t)$ [6] involves modified Bessel functions of the first and second kind and can be deduced from (2.9) below, but there is a simple expression for $f(t)$ for times that extend to $\mathcal{O}(k^{-1})$,

$$f(t) = \frac{\beta}{\pi} \tanh\left[\frac{2\pi}{k\beta}\mathcal{S}(t)\right] \quad \text{where } \mathcal{S}(t) = \left(1 - e^{-kt/2}\right)\theta(t). \tag{2.7}$$

This approximation will be adequate for our analysis here, but longer time scales of order $k^{-1}|\log \beta k|$ require the full Bessel functions.

The coordinates $x^\pm$ do not extend behind the horizon and so it will prove useful to define new coordinates $w^\pm$ related to $x^\pm$ by a Möbius transformation[7]

$$x^\pm = \pm\frac{\beta}{\pi} \cdot \frac{w^\pm \mp 1}{w^\pm \pm 1}, \tag{2.8}$$

which cover the region inside the horizon of the newly created black hole, $w^- > 0$. The horizon of the original extremal black hole in front of the shockwave is at $w^- = 1$, while, behind the shockwave the horizon is at $w^- = 0$. It is important that the vacuum state does not change under a Möbius, or $\text{SL}(2,\mathbb{R})$ transformation.

The new coordinates are related to the bath coordinates by an associated function $w^\pm = \pm\hat{f}(y^\pm)^{\pm 1}$. The function $\hat{f}$ is expressed in terms of modified Bessel functions of the 1st and 2nd kind [6],

$$\hat{f}(t) = \frac{e^{4\pi/(\beta k)}}{\pi} \frac{K_0(z)}{I_0(z)}, \quad z = \frac{2\pi}{\beta k} e^{-kt/2}. \tag{2.9}$$

However, in the small k regime with kt fixed, it takes a simple form which we write through the shockwave, and for times up to $\mathcal{O}(k^{-1})$, as

$$\log \hat{f}(t) = \theta(-t)\log\frac{1 + \pi t/\beta}{1 - \pi t/\beta} + \frac{4\pi}{\beta k}\mathcal{S}(t). \tag{2.10}$$

The expression for $t < 0$ yields $x^\pm = t$ via the Möbius transform (2.8).

---

[6] Whilst still requiring $S_0 \gg 1$ to suppress higher terms in the genus expansion of JT gravity.
[7] In an equilibrium situation at temperature β^{-1}, $w^\pm = \pm\exp(\pm 2\pi y^\pm/\beta)$ where $y^\pm$ cover the black hole patch outside the horizon.

In what follows, we will need the small k limit, with kt and β fixed, of various quantities that depend on f and $\hat{f}$ evaluated at various times. We will use the notation $\hat{f}_i = \hat{f}(t - \tau_i)$, etc. The leading behaviour we want is always $\mathcal{O}(k^{-1})$. For example, consider $\log \hat{f}_i$, whose leading behaviour is

$$\log \hat{f}_i \to \frac{4\pi}{\beta k} \mathcal{S}(t - \tau_i). \tag{2.11}$$

So at this order, $\log \hat{f}_i$ switches on abruptly at $t = \tau_i$. Another important example is the quantity $\log(\hat{f}_j^{-1} - \hat{f}_i^{-1})$ with, say, $\tau_i < \tau_j$. In this case, the contribution at order k^{-1} kicks in when $t > \tau_j$, but independent of τ_i,

$$\log(\hat{f}_j^{-1} - \hat{f}_i^{-1}) \to -\frac{4\pi}{\beta k} \mathcal{S}(t - \tau_j). \tag{2.12}$$

Similarly $\log \hat{f}_i' \to \frac{4\pi}{\beta k} \mathcal{S}(t - \tau_i)$. Finally, we can deal with expressions that depend on f rather than $\hat{f}$ by effectively replacing $f_i \to 1 - 2\hat{f}_i^{-1}$, which follows from the Möbius transformation that relate the two, in the small k limit.

3. Entropy of arbitrary intervals

As the excited black hole state evaporates and relaxes, we want to evaluate the von Neumann entropy in the Hawking radiation collected by a certain number of disjoint intervals in the bath.

3.1. No-island saddle

We consider a union $A \equiv \cup_{j=1}^N A_j$ of N disjoint spatial intervals in the bath

$$A_j = [b_{2j-1}, b_{2j}] \quad \text{with } b_j < b_{j+1}, \ j = 1, 2 \ldots N, \tag{3.1}$$

with Minkowski coordinates for each endpoint p_j: $(y^\pm = t \pm b_j)$. We will be interested in timescales t of order k^{-1} and locations $\{b_j\}$ which are also $\mathcal{O}(k^{-1})$. The entanglement entropy of the union of these intervals, in a theory of $2c$ free fermions in the vacuum state in a non-trivial frame,

$$ds^2 = -\Omega^{-2} \, d\xi^+ \, d\xi^-, \tag{3.2}$$

is given by,

$$S_{\text{QFT}} = -\frac{c}{6} \sum_{i<j} (-1)^{j-i} \log \sigma_{ij} - \frac{c}{6} \sum_j \log \Omega_j + N S_{\text{UV}}, \tag{3.3}$$

where $\sigma_{ij} = -(\xi_i^+ - \xi_j^+)(\xi_i^- - \xi_j^-)$ and $S_{\text{UV}} = -\frac{c}{6} \log \varepsilon_{\text{UV}}$ contains the UV cut off, while Ω_j are the values of the conformal factor at each endpoint. The result follows from the expression for the entropy of several intervals of free fermions in the Minkowski vacuum [31] augmented with a conformal transformation to the non-trivial frame [6].

In the present case, the in-going modes are in the vacuum state of the bath coordinate y^+ whereas the out-going modes, containing the Hawking radiation, are in the vacuum state of the AdS coordinate w^- (or equivalently x^-). Hence, the out-going modes—the Hawking radiation—once they are in the bath are manifestly not in the Minkowski vacuum.

The QFT entropy, what we will call the 'no-island saddle', is computed by using the frame (y^+, w^-) and takes the form

$$S_{\text{no island}} \equiv S_{\text{QFT}} = -\frac{c}{6}\sum_{i<j}(-1)^{j-i}\,\log(\hat{f}_j^{-1} - \hat{f}_i^{-1})(y_j^+ - y_i^+) + \frac{c}{12}\sum_j \log\frac{\hat{f}_j^2}{\hat{f}'_j} + \cdots \quad (3.4)$$

The second set of terms arise from the conformal factors at $\{p_j\}$ from changing coordinates from y^- to w^-. Now we take the small k limit, applying the rules that we established earlier. Note that the y^+ coordinates do not contribute at leading order in k^{-1}:

$$S_{\text{no island}} = 2S_{\text{BH}}\sum_{j=1}^{2N}(-1)^{j-1}\mathcal{S}_{b_j} + NS_{\text{UV}}, \quad (3.5)$$

where we used the shorthand $\mathcal{S}_b \equiv \mathcal{S}(t-b)$. So in the slow evaporation limit, the entropy takes a simple form in terms of the elementary function $\mathcal{S}(t)$ defined in (2.7). The final form above is consistent with causality. The wavefront of the Hawking radiation as it enters the bath, is along $y^- = 0$. It only reaches an endpoint b of an interval at $t = b$ which is reflected in the fact that $\mathcal{S}(t-b)$ is proportional to the step function $\theta(t-b)$.

3.2. Island saddles

The entropy of regions $A = \cup_j A_j$ in the bath that we have just established, can be computed semi-classically using steepest-descent functional integral techniques via the replica method [32, 33]. The recent step forward, is the realization that in the gravitational context there can be new saddles that involve non-trivial replica geometries, the *replica wormholes* [1, 2]. The new saddles, or instantons, can be computed by a recipe as follows: if we are computing an entropy of a region in the bath A, then the replica-wormhole saddles require us to define a generalized entropy by adding to the QFT entropy, the contribution of new regions I, the island(s), whose endpoints $p_{\hat{a}}$ are the *quantum extremal surfaces*[8], and also the value of the dilaton at the QESs. Then one extremizes over the positions of the QESs:

$$S_I(A) = \underset{\{p_{\hat{a}}\}}{\text{ext}}\left\{\sum_{\hat{a}}\frac{\phi(p_{\hat{a}})}{4G_N} + S_{\text{QFT}}(A \cup I)\right\}. \quad (3.6)$$

Finally, it is the saddle with the smallest entropy amongst the saddles that dominates,

$$S_A = \min_I S_I(A), \quad (3.7)$$

including the no-island saddle. The recipe of this kind of form first appeared in holographic contexts [34–37] but the replica wormholes mean that the result can be derived from conventional field theory methods without any recourse to holography. Note that usually, the dilaton terms cost a lot of entropy and so it is only in very special situations that an island saddle dominates over the no-island saddle.

3.3. Two-QES saddles: one behind shockwave

There are various kinds of possible island saddles that could contribute. We will assume here that the only important ones have two QES $p_{\hat{a}}, \hat{a} = 1, 2$. Consider the generalized entropy—the

---

[8] Strictly speaking the island is the causal domain of a Cauchy slice joining the QES. Note that QESs are just points in 2d gravity, but in higher dimensions they are genuine surfaces.

terms in the curly brackets in (3.6)—in $x^\pm$ coordinates

$$S_{\text{gen.}}(x_{\hat{a}}^\pm) = 2S_0 + \frac{c}{6}\log(x_{\hat{1}}^+ - x_{\hat{2}}^+)(x_{\hat{2}}^- - x_{\hat{1}}^-)$$

$$+ \frac{c}{6}\sum_{\hat{a}=1}^{2}\left\{\frac{1}{k}\cdot\frac{1}{x_{\hat{a}}^- - x_{\hat{a}}^+} + \log\frac{2}{x_{\hat{a}}^- - x_{\hat{a}}^+} - \sum_{j=1}^{2N}(-1)^{\hat{a}-j}\right.$$

$$\left.\times\,\log(t + b_j - x_{\hat{a}}^+)(x_{\hat{a}}^- - f_j)\right\} + S_{\text{no island}}. \tag{3.8}$$

It is clear from this that extrema would have $x_{\hat{a}}^\pm = \mathcal{O}(k^{-1})$ and the entropy is dominated at $\mathcal{O}(k^{-1})$ by the no-island contribution. Actually, closer analysis fails to find a physically acceptable solution and so we will discard this possibility anyway.

Now consider the case with one QES in front and one behind the shockwave, the case illustrated in figure 1. The QES in front $p_{\hat{1}}$ will have coordinates $x_{\hat{1}}^\pm = \mathcal{O}(k^{-1})$, i.e. $w_{\hat{1}}^\pm = \mp 1$, to leading order, and so our task is to find the position of the second QES $p_{\hat{2}}$ behind the shockwave. The ingredients we need, include the conformal factor in the AdS region in the (y^+, w^-) frame,

$$\Omega^{-2} = \frac{4\hat{f}'(y^+)}{(1 + w^+w^-)^2}, \tag{3.9}$$

and the dilaton, determined by the master function $\hat{f}$ [11]:

$$\phi = \phi_0 + 2\phi_r\left\{\frac{\hat{f}''(y^+)}{2\hat{f}'(y^+)} - \frac{w^-\hat{f}'(y^+)}{1 + w^-\hat{f}(y^+)}\right\}. \tag{3.10}$$

In the same frame, it is useful to change variable from y^+ back to $w^+ = \hat{f}(y^+)$. Within our approximation (2.10), we then have

$$\phi = \phi_0 + \frac{2\pi\phi_r}{\beta}\cdot\frac{1 - w^+w^-}{1 + w^+w^-}\left(1 - \frac{\beta k}{4\pi}\log\,w^+\right) - \frac{k\phi_r}{2},$$

$$\Omega^{-2} = \frac{4}{(1 + w^+w^-)^2}\cdot\frac{2\pi w^+}{\beta}\left(1 - \frac{\beta k}{4\pi}\log\,w^+\right). \tag{3.11}$$

Putting the ingredients together, the generalized entropy needed to determine the position of the second QES, is

$$S_{\text{gen.}}(w_{\hat{2}}^\pm) = \frac{\pi c}{6\beta k}\cdot\frac{1 - w_{\hat{2}}^+w_{\hat{2}}^-}{1 + w_{\hat{2}}^+w_{\hat{2}}^-}\left(1 - \frac{\beta k}{4\pi}\log\,w_{\hat{2}}^+\right)$$

$$+ \frac{c}{12}\log\frac{w_{\hat{2}}^+(1 - \frac{\beta k}{4\pi}\log\,w_{\hat{2}}^+)}{(1 + w_{\hat{2}}^+w_{\hat{2}}^-)^2} + \frac{c}{6}\log(w_{\hat{1}}^- - w_{\hat{2}}^-)(y_{\hat{2}}^+ - x_{\hat{1}}^+)$$

$$- \frac{c}{6}\sum_j(-1)^j\log(w_{\hat{2}}^- + \hat{f}_j^{-1})(t + b_j - y_{\hat{2}}^+) + S_{\text{no island}} + \cdots. \tag{3.12}$$

We have not shown some constants that are sub-leading in k.

Extremizing over the coordinates $w_{\hat{2}}^\pm$ of the second QES, leads to a pair of rather complicated equations. However, only a small number of terms actually matter in the small k limit. In

particular, the solutions will have $w_2^+ w_2^- = \mathcal{O}(k)$ and this allows us to replace $1 + w_2^+ w_2^- \to 1$. In addition, there are terms contributing to the saddle point equation from varying with respect to w_2^- of the form

$$\frac{\beta}{2\pi w_2^+} \cdot \frac{1}{y_2^+ - t - b_j}, \qquad \frac{\beta}{2\pi w_2^+} \cdot \frac{1}{y_2^+ - x_1^+}, \tag{3.13}$$

Since we take t, b_j to be $\mathcal{O}(k^{-1})$, and once we have the solution, y_2^+ is also $\mathcal{O}(k^{-1})$, these terms are suppressed by a factor of k compared to leading terms in the saddle point equations.

The stripped down equations that determine the leading order behaviour can be written compactly in terms of the variables ω_2^+ and w_2^- where

$$\omega_2^+ = w_2^+ \left(1 - \frac{\beta k}{4\pi} \ln\ w_2^+\right) = w_2^+ \left(1 - \mathcal{S}(y_2^+)\right). \tag{3.14}$$

It is useful to recall the definition of $\mathcal{S}$ and note that $\omega_2^+ = w_2^+ e^{-k y_2^+/2}$. For the saddle point to be behind the shockwave, we need $w_2^+ > 0$, or equivalently $\omega_2^+ > 0$. Importantly, on the time and length scales of our interest, the pre-factor $(1 - \mathcal{S})$ is $\mathcal{O}(k^0)$. We find,

$$\frac{2\pi}{\beta k} \omega_2^+ - \frac{1}{w_2^- - w_1^-} + \sum_{j=1}^{2N} \frac{(-1)^j}{w_2^- + \hat{f}_j^{-1}} = 0,$$

$$\frac{2\pi}{\beta k} w_2^- - \frac{1}{4\,\omega_2^+} = 0. \tag{3.15}$$

The second of these two equations ensures that $w_2^+ w_2^- = \mathcal{O}(k)$, consistent with the approximation.

We can now insert the position of the first QES, $w_1^- = 1$ into the above. If we write $w_2^- = 1/4p$, then $\omega_2^+ = \beta k p/2\pi$ and p must satisfy

$$1 + \frac{4}{4p - 1} + 4 \sum_{j=1}^{2N} \frac{(-1)^j \hat{f}_j}{4p + \hat{f}_j} = 0. \tag{3.16}$$

The solution of this equation is made straightforward by the extreme behaviour of the functions $\hat{f}_j$ in the small k limit. At leading order in k, it vanishes for $t < b_j$, but then increases exponentially for $t > b_j$ in such a way that $\hat{f}_j \gg \hat{f}_i$, if $b_j < b_i$. Hence, one can spot $2N$ distinct solutions labelled by $\alpha = 1, 2, \ldots, 2N$, for which $p = c_\alpha \hat{f}_\alpha$, i.e.

$$\omega_2^+ = \frac{\beta k c_\alpha}{2\pi} \hat{f}_\alpha, \qquad w_2^- = \frac{1}{4 c_\alpha \hat{f}_\alpha}, \tag{3.17}$$

where $c_\alpha = \frac{3}{4}$, for α odd, and $c_\alpha = \frac{1}{12}$, for α even. Working to the leading order in the small k approximation, the coordinate $w_2^+ = \omega_2^+ e^{k y_2^+/2}$ for the QES behind the shockwave is then simply,

$$w_2^+ = \frac{\beta k c_\alpha}{2\pi} \hat{f}_\alpha e^{k y_\alpha^-/2}. \tag{3.18}$$

We can also write (3.17) in terms of the w^- coordinate of the corresponding point in the bath $w_\alpha^- = -1/\hat{f}_\alpha$

$$\omega_{\hat{2}}^+ = -\frac{\beta k c_\alpha}{2\pi} \cdot \frac{1}{w_\alpha^-}, \qquad w_{\hat{2}}^- = -\frac{w_\alpha^-}{4c_\alpha}, \tag{3.19}$$

Note that the saddles are only consistent if $w_{\hat{2}}^+ > 0$, ensuring that the QES is behind the shockwave: see figure 1. This means that they only appear when $t > b_\alpha$. In addition, we have $w_{\hat{2}}^+ < w_1^+$ which is necessary because the QES $p_{\hat{2}}$ must be space-like with respect to the points in the bath. All these saddles have $w_{\hat{2}}^- > 0$, and so the second QES is inside the horizon of the black hole created by the shockwave at least in our window of approximation, i.e. for times t of $\mathcal{O}(k^{-1})$. For much later times, the QES pops outside the horizon as the black hole relaxes back to the extremal black hole [11]. We will not be concerned with this very long time regime in this paper.

In order to calculate the entropy of the island saddle with the solution (3.17), the relevant terms in (3.12), that contribute at leading order, are

$$\frac{\pi c}{6\beta k} \cdot \frac{1 - w_{\hat{2}}^+ w_{\hat{2}}^-}{1 + w_{\hat{2}}^+ w_{\hat{2}}^-} \left(1 - \frac{\beta k}{4\pi} \log\, w_{\hat{2}}^+ \right) + \frac{c}{12} \log\, w_{\hat{2}}^+ \to S_{\text{BH}} \left(1 + \mathcal{S}_{b_\alpha} \right),$$

$$-\frac{c}{6} \sum_j (-1)^j \, \log(w_{\hat{2}}^- + \hat{f}_j^{-1}) \to 4S_{\text{BH}} \sum_j (-1)^j \mathcal{S}_{b_{\max(\alpha,j)}}. \tag{3.20}$$

The entropy only depends on $\log\, w_{\hat{2}}^+$ at the leading order in k wherein $\log\, w_{\hat{2}}^+ \simeq \log\, \omega_{\hat{2}}^+ = \mathcal{O}(k^{-1})$. Hence, the entropy of the α-labelled island saddle, $t > b_\alpha$, is

$$S_{\text{island}(\alpha)} = S_{\text{no island}} + S_{\text{BH}} \left\{ 1 + \mathcal{S}_{b_\alpha} + 4\sum_{j=1}^{2N} (-1)^j \mathcal{S}_{b_{\max(j,\alpha)}} \right\}. \tag{3.21}$$

If the extremal entropy is not negligible, we must add $2S_0$ to the island saddle entropies, but for the present purpose we will mostly ignore it. Note that the islands do not lead to additional UV divergences because these are absorbed into a renormalization of ϕ_0. The island saddles with α even, never have the lowest entropy and can therefore be ignored. This follows from the inequalities,

$$S_{\text{island}(2\alpha)} - S_{\text{island}(2\alpha\pm1)} = (-1 \pm 2)S_{\text{BH}} \, e^{-kt/2} (e^{kb_{2\alpha\pm1}/2} - e^{kb_{2\alpha}/2}) > 0. \tag{3.22}$$

3.4. Both QES behind the shockwave

Now let us consider the possibility that both QES are behind the shockwave. The generalized entropy is

$$S_{\text{gen.}}(w_{\hat{a}}^\pm) = \sum_{\hat{a}} \left(\frac{\phi(p_{\hat{a}})}{4G_N} - \frac{c}{6} \log\, \Omega_{\hat{a}} - \sum_j (-1)^{\hat{a}-j} \, \log\, \sigma_{\hat{a}j} \right)$$

$$- \sum_{\hat{a}<\hat{b}} (-1)^{\hat{a}-\hat{b}} \, \log\, \sigma_{\hat{a}\hat{b}} + S_{\text{no island}}, \tag{3.23}$$

where, for each QES, the dilaton and conformal factor are given in (3.12). In the above equation, the spacetime intervals are

$$\sigma_{\hat{a}\hat{b}} = -(w_{\hat{a}}^- - w_{\hat{b}}^-)(y_{\hat{a}}^+ - y_{\hat{b}}^+), \qquad \sigma_{\hat{a}j} = -(w_{\hat{a}}^- + \hat{f}_j^{-1})(y_{\hat{a}}^+ - t - b_j). \quad (3.24)$$

We are assuming that the order of the points on a Cauchy surface, from left to right, is $(p_{\hat{1}}, p_{\hat{2}}, p_1, p_2, \ldots, p_{2N})$.

The position of the second QES $p_{\hat{2}}$ is again determined by equation (3.15), and now the first QES $p_{\hat{1}}$ satisfies the same equation with the replacement $p_{\hat{2}} \leftrightarrow p_{\hat{1}}$,

$$\begin{aligned}
\frac{2\pi}{\beta k}\omega_{\hat{1}}^+ - \frac{1}{w_{\hat{1}}^- - w_{\hat{2}}^-} - \sum_{j=1}^{2N} \frac{(-1)^j}{w_{\hat{1}}^- + \hat{f}_j^{-1}} &= 0, \\
\frac{2\pi}{\beta k}w_{\hat{1}}^- - \frac{1}{4\,\omega_{\hat{1}}^+} &= 0.
\end{aligned} \quad (3.25)$$

Again we have self-consistently discarded terms that are higher order in k. Each QES has $2N$ possible saddle points labelled as $\alpha(\hat{a})$ where $\hat{a} = 1, 2$. The solutions generalize (3.17) with

$$w_{\hat{a}}^+ = \frac{\beta k c_{\hat{a}}}{2\pi}\hat{f}_{\alpha(\hat{a})}\, e^{\frac{1}{2}ky_{\alpha(\hat{a})}^-}, \qquad w_{\hat{a}}^- = \frac{1}{4c_{\hat{a}}\hat{f}_{\alpha(\hat{a})}}. \quad (3.26)$$

Here $c_{\hat{a}} = \frac{3}{4}$ for $\alpha(\hat{1})$ even or $\alpha(\hat{2})$ odd and $c_{\hat{a}} = \frac{1}{12}$ for $\alpha(\hat{1})$ odd or $\alpha(\hat{2})$ even. Since $w_{\hat{1}}^+ < w_{\hat{2}}^+$, we must have[9] $\alpha(\hat{1}) > \alpha(\hat{2})$.

We note that there is a term involving $(y_{\hat{1}}^+ - y_{\hat{2}}^+)$ in the second equation in (3.15) and in (3.23) which contributes in the special case that $\alpha(\hat{1}) = \alpha(\hat{2})$. But this case yields an entropy that is always greater than the no-island entropy:

$$S_{\text{island}(\alpha,\alpha)} = S_{\text{no island}} + 2S_{\text{BH}}\left(1 - \mathcal{S}_{b_\alpha}\right) > S_{\text{no island}} \quad (3.27)$$

and so is never dominant. Consequently from now on, we will assume that $\alpha(\hat{2}) \neq \alpha(\hat{1})$.

Since we are taking both QESs to be behind the shockwave, the saddle points above only appear when $t > b_{\alpha(\hat{1})}$ and they have entropy given by evaluating (3.12) in the small k regime, yielding,

$$S_{\text{island}(\alpha(\hat{1}),\alpha(\hat{2}))} = S_{\text{no island}} + S_{\text{BH}}\left\{ 2 + \sum_{\hat{a}=1}^{2}\left(\mathcal{S}_{b_{\alpha(\hat{a})}} + 4\sum_{j=1}^{2N}(-1)^{j-\hat{a}}\mathcal{S}_{b_{\max(j,\alpha(\hat{a}))}}\right) - 4\mathcal{S}_{b_{\alpha(\hat{1})}}\right\}. \quad (3.28)$$

There is a zoo of possible saddles that can compete for the minimal entropy. However, many of the saddles can never have minimal entropy, given that $1 \geqslant \mathcal{S}_{b_i} \geqslant \mathcal{S}_{b_j}$, for $i < j$. The following is a list of saddles for one and two intervals which can have minimal entropy. We write them

---

[9] $\omega_{\hat{a}}^+$ and $w_{\hat{a}}^+$ only differ by order one terms in the exponent, and therefore on the time scales of interest, $\log \omega_{\hat{1}}^+ < \log \omega_{\hat{2}}^+ \Rightarrow \log w_{\hat{1}}^+ < \log w_{\hat{2}}^+$.

in shorthand as a vector $(c_0, c_1, \ldots, c_{2N})$,

$$S = S_{\text{BH}} \left(c_0 + \sum_{j=1}^{2N} c_j \mathcal{S}_j \right). \tag{3.29}$$

Using this notation we explicitly list S_{noisland}, the island entropy with one QES behind the shockwave $S_{\text{island}(\alpha)}$, and the entropy $S_{\text{island}(\alpha(\hat{1}),\alpha(\hat{2}))}$ with both QESs behind the shockwave:

$$S_{\text{no island}} = (0, 2, -2), \qquad S_{\text{island}(1)} = (1, -1, 2), \qquad S_{\text{island}(21)} = (2, -1, -1). \tag{3.30}$$

Then, for two intervals we have,

$$\begin{aligned}
S_{\text{no island}} &= (0, 2, -2, 2, -2), & S_{\text{island}(1)} &= (1, -1, 2, -2, 2), \\
S_{\text{island}(3)} &= (1, 2, -2, -1, 2), & S_{\text{island}(21)} &= (2, -1, -1, 2, -2), \\
S_{\text{island}(41)} &= (2, -1, 2, -2, -1), & S_{\text{island}(43)} &= (2, 2, -2, -1, -1).
\end{aligned} \tag{3.31}$$

Note that in the expressions above, we have left the cut off term implicit.

3.5. Position of the QES and scrambling time

It is important to know when an in-going null ray beginning on the boundary at time t_0 lies in the island of a saddle at time t. Given the coordinates of the QESs in (3.26), the condition is

$$\frac{\beta k c_{\alpha(\hat{1})}}{2\pi} e^{\frac{1}{2}ky^-_{\alpha(\hat{1})}} \hat{f}_{\alpha(\hat{1})} < \hat{f}(t_0) < \frac{\beta k c_{\alpha(\hat{2})}}{2\pi} e^{\frac{1}{2}ky^-_{\alpha(\hat{2})}} \hat{f}_{\alpha(\hat{2})}, \tag{3.32}$$

where the lower bound is only important if both QES are behind the shockwave. The bounds here, using our approximations, are

$$t_0 = t - b_{\alpha(\hat{a})} - e^{k(t-b_{\alpha(\hat{a})})/2} \frac{\beta}{2\pi} \log \left(\frac{2\pi}{\beta k c_{\alpha(\hat{a})}} e^{-k(t-b_{\alpha(\hat{a})})/2} \right). \tag{3.33}$$

The final term here is an expression of the *scrambling time*, the time it takes to recover information dropped into an old black hole as we shall verify later. This term is sub-leading in the small k limit, where t_0 and the b_j are order k^{-1}. Interestingly, if we ignore the small correction from the scrambling time and partner the in-going ray at $t_0 = t - b_{\alpha(\hat{a})}$ on the boundary with an out-going ray, then the latter hits the branch point $p_{\alpha(\hat{a})}$ in the bath at time t just as the in-going one hits the QES $p_{\hat{a}}$, as shown in figure 2.

4. Page curve and early/late correlations

The Page curve is simply the entropy of a single interval $[0, t]$ in the bath at time t that collects all the Hawking radiation emitted during the life of the evaporating black hole up to time t, i.e. in the temporal interval $\langle 0, t \rangle$.[10] Note that the point p_2 lies on the wave front of the Hawking

---

[10] We use angle brackets to distinguish a temporal interval from a spatial one and we will specify both to avoid any confusion. Importantly, when we write $[0, t]$ we mean the limit $\epsilon \to 0$ of $[\epsilon, t]$, to emphasize that in the holographic interpretation the boundary system is *not* included.
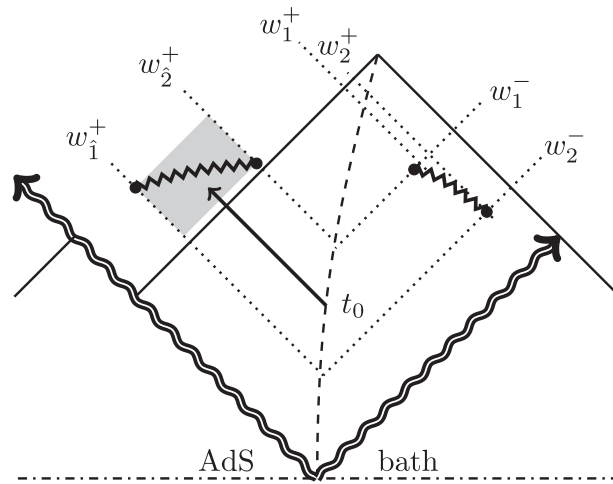
Figure 2. The w^+ coordinates of the QESs determine when a null ray from the boundary is in the island of an interval in the bath. The shaded area is the island-(21) of a single interval in the bath at time t. Note, also, the relation between the w^- coordinates branch points in the bath and the w^+ coordinates of the QESs.

radiation. The curve involves a competition between the no-island and island-(1) saddles:

$$S_{\text{Page curve}}(t) = S_{\text{BH}} \ \min\left(2\mathcal{S}(t), 1 - \mathcal{S}(t)\right)$$

$$= S_{\text{BH}} \ \min\left(2 - 2 \ e^{-kt/2}, e^{-kt/2}\right). \tag{4.1}$$

The Page time occurs when the no-island ceases to dominate and the island saddle takes over. This occurs at

$$t_{\text{Page}} = \frac{2}{k} \log \frac{3}{2}. \tag{4.2}$$

If we do not neglect the extremal entropy and write $S_0 = \xi S_{\text{BH}}$, then the Page time is increased to $t_{\text{Page}} = \frac{2}{k} \log \frac{3}{2(1-\xi)}$. Clearly, if $\xi > 1$ then our approximation regime breaks down and full Bessel function expression for $\hat{f}$ will be needed. In the following, for simplicity, we will suppose that ξ is small and can be neglected.

4.1. Mutual information

We can compute the mutual information between the early and late radiation emitted in time intervals $\langle 0, t_{\text{Page}} \rangle$ and $\langle t_{\text{Page}}, t \rangle$, for $t > t_{\text{Page}}$. In order to do this, one collects the two subsets in spatial intervals $B = [t - t_{\text{Page}}, t]$ and $A = [0, t - t_{\text{Page}}]$, respectively, at time t: see figure 3. In order for the result to be UV safe, we suppose the two intervals are slightly separated by more than the UV cut off. The mutual information is

$$I_{A,B} = S_A + S_B - S_{AB}. \tag{4.3}$$

It is easy to see that B is at the island-(1)/no-island threshold, while AB is always in its island-(1) saddle. The interval A has a transition from its no-island to island saddle at a time
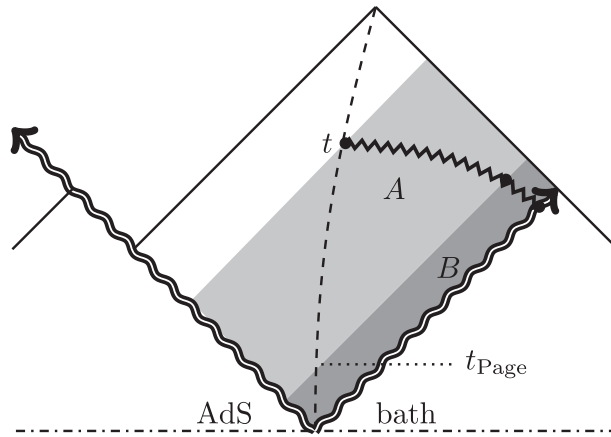
Figure 3. A scenario to measure the mutual information of the early and late Hawking radiation emitted in time intervals $B = \langle 0, t_{\text{Page}} \rangle$ and $A = \langle t_{\text{Page}}, t \rangle$ by collecting the radiation in appropriate spatial intervals on a Cauchy surface.

$t = \frac{2}{k} \log \frac{9}{2}$:

$$S_A = S_{\text{BH}} \min \left(\frac{4}{3} - 2\,e^{-kt/2}, \frac{2}{3} + e^{-kt/2} \right),$$

$$S_B = \frac{2}{3} S_{\text{BH}}, \qquad S_{AB} = S_{\text{BH}}\,e^{-kt/2}. \tag{4.4}$$

Hence, the mutual information increases as t increases from t_{Page}, but then plateaus for $t > \frac{2}{k} \log \frac{9}{2}$:

$$I_{A,B} = S_{\text{BH}} \min \left(2 - 3\,e^{-kt/2}, \frac{4}{3} \right). \tag{4.5}$$

4.2. Quantum correlations

The fact that the early and late modes have non-vanishing mutual information means that they are correlated but this does not discriminate between classical and quantum correlations. One way to measure genuine quantum correlations, is via the behaviour of the conditional entropies [38]

$$S_{A|B} = S_{AB} - S_B, \qquad S_{B|A} = S_{AB} - S_A. \tag{4.6}$$

These must be non-negative in a classical system and so their negativity is a measure of quantum correlations. In the present case,

$$S_{A|B} = S_{\text{BH}} \left(e^{-kt/2} - \frac{2}{3} \right), \qquad S_{B|A} = -S_{\text{BH}} \min \left(\frac{2}{3}, \frac{4}{3} - 3\,e^{-kt/2} \right). \tag{4.7}$$

Recalling that $t > t_{\text{Page}}$, the former is always negative and the latter becomes negative for $t > \frac{2}{k} \log \frac{9}{2}$.

We can consider the same correlation measures on other subsets of the radiation. For example in figure 4, we show the conditional entropies for two subsets of the radiation $\langle 0, t \rangle$ and $\langle t, 4t_{\text{Page}} \rangle$ with varying t. The existence of genuine quantum correlation is clear.
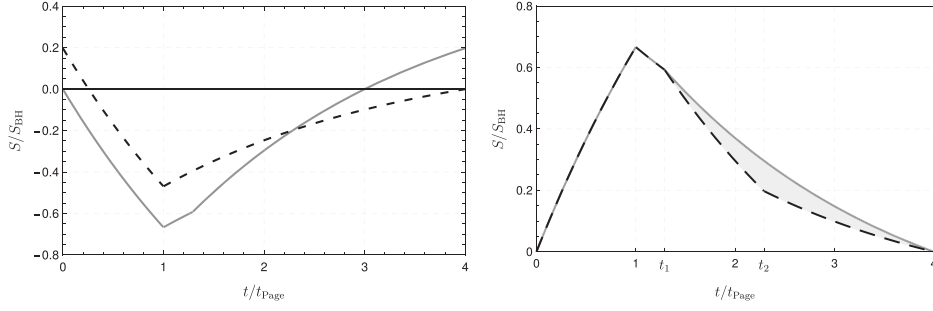
Figure 4. Left: A plot of $S_{A|B}$ (dashed) and $S_{B|A}$ (solid) for the two temporal subsets $B = \langle 0, t \rangle$ and $A = \langle t, 4t_{\text{Page}} \rangle$. Notice that at least one of them is negative at each t indicating the presence of quantum correlations. Right: for the same regions a plot of the upper and lower bounds on the squashed entanglement (assuming the UV cut off term is sufficiently small). Thus proves the existence of entanglement between A and B.

Measuring entanglement between two subsystems A and B in quantum mechanics is not simple or obvious when the state on AB is not pure. An excellent measure of quantum correlations is provided by the *squashed entanglement* [39–41] (see also [42, 44]) defined as

$$E_{A,B}^{\text{sq.}} = \frac{1}{2} \min_C I_{A,B|C}, \tag{4.8}$$

involving the *conditional mutual information*

$$I_{A,B|C} = I_{A,BC} - I_{A,C}. \tag{4.9}$$

Here, C is *any* additional quantum system appended to A and B. Unfortunately, this latter feature renders it impractical to calculate the squashed entanglement. However, it can be bounded both from above and below [41] by quantities that are calculable:

$$\max(0, -S_{A|B}, -S_{B|A}) \leqslant E_{A,B}^{\text{sq.}} \leqslant \frac{1}{2} I_{A,B}. \tag{4.10}$$

Notice that the lower bound involves minus the conditional entropies, so a non-trivial bound occurs when one or both of the conditional entropies are negative, indicating quantum entanglement. In the QFT setting, the lower bound can be made UV safe, if A and B (but not C) are taken to be adjacent.

Let us consider two subsets of radiation $\langle 0, t \rangle$ and $\langle t, T \rangle$ collected in bath regions $A = [0, T - t]$ and $B = [T - t, T]$ at time T. One can compute the upper and lower bounds of the squashed entanglement, $E_{\gtrless}^{\text{sq.}}$, as in (4.10). There are four temporal regions:

$$
\begin{aligned}
t \leqslant t_{\text{Page}} \qquad & E_{<}^{\text{sq.}} = E_{>}^{\text{sq.}} = 2\mathcal{S}(t), \\[4pt]
t_{\text{Page}} < t \leqslant t_1 \qquad & E_{<}^{\text{sq.}}(-S_{B|A}) = E_{>}^{\text{sq.}} = 1 - \mathcal{S}(t), \\[4pt]
t_1 < t \leqslant t_2 \qquad &
\begin{cases}
E_{<}^{\text{sq.}}(-S_{B|A}) = -1 + 3\mathcal{S}(T) - 2\mathcal{S}(t), \\
E_{>}^{\text{sq.}} = \dfrac{3}{2}(\mathcal{S}(T) - \mathcal{S}(t)),
\end{cases} \\[4pt]
t > t_2 \qquad &
\begin{cases}
E_{<}^{\text{sq.}}(-S_{A|B}) = \mathcal{S}(T) - \mathcal{S}(t), \\
E_{>}^{\text{sq.}} = \dfrac{3}{2}(\mathcal{S}(T) - \mathcal{S}(t)),
\end{cases}
\end{aligned}
\tag{4.11}
$$

16
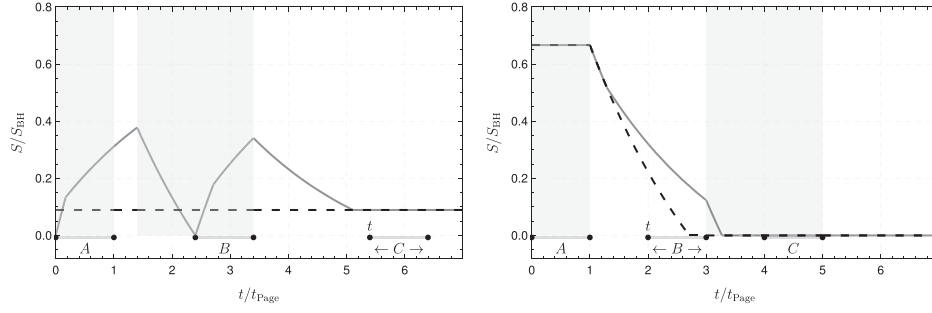
Figure 5. These plots show the mutual information $I_{A,B}$ (dotted) and conditional mutual information $I_{A,B|C}$ (continuous) for three intervals A, B, C in the Hawking radiation collected at a fixed time $8t_{\text{Page}}$. Left: with A and B fixed with C moving and Right: with A and C fixed with B moving. The shaded regions indicates where the intervals overlap. Notice that we get $I_{A,B|C} = 0$ when $C = A$ or B, as expected from its definition. The non-negativity of $I_{A,B}$ and $I_{A,B|C}$ are checks of sub-additivity and strong sub-additivity, respectively.

where $t_1 = T - \frac{\log 3}{\log 3/2} t_{\text{Page}}$ and $t_2 = T - \frac{\log 2}{\log 3/2} t_{\text{Page}}$. Figure 4 shows the upper and lower bounds for our example with $T = 4t_{\text{Page}}$.

Even though we cannot calculate the squashed entanglement directly, we can gain some intuition by choosing C in (4.8) to be another subset of the Hawking radiation. In figure 5 we illustrate that, when the interval C is distinct, i.e. has no overlap with A and C, that

$$I_{A,B|C} \geqslant I_{A,B}. \tag{4.12}$$

So, even though we cannot claim that $E^{\text{sq.}}_{A,B} = \frac{1}{2} I_{A,B}$, this is an indication that the mutual information we have calculated is a good measure of entanglement since we managed to 'squash away' some of the non-quantum correlations by choosing a suitable C of the Hawking radiation. Moreover, since the mutual information is an upper bound for the squashed entanglement, this is another indication that the correlations amongst subsets of the Hawking radiation are due to entanglement and are not just classical.

There are other interesting measures of entanglement like the reflected entropy that can also be measured in context of black holes in JT gravity [43].

4.3. Entropy consistency conditions

The von Neumann entropies of entanglement must satisfy various consistency conditions in a generic quantum system, including sub-additivity and the Araki–Lieb inequality. Firstly, if we assume that the UV cut-off is sub-leading compared with the order k^{-1} contributions to the entropy, then one can verify that the Araki–Lieb triangle inequality,

$$S_{AB} \geqslant |S_A - S_B|, \tag{4.13}$$

is satisfied by our subset of modes $\langle 0, t_{\text{Page}} \rangle$ and $\langle t_{\text{Page}}, t \rangle$. We can also make the same check for other subsets; for example in figure 6, we show S_{AB}, $S_A + S_B$ and $|S_A - S_B|$ for subsets $\langle 0, t - \frac{1}{2} t_{\text{Page}} \rangle$ and $\langle t, 5t_{\text{Page}} \rangle$ as a function of t.

The next consistency condition is sub-additivity; namely the non-negativity of the mutual information $I_{A,B} \geqslant 0$. This is clearly satisfied for our subsets $\langle 0, t_{\text{Page}} \rangle$ and $\langle t_{\text{Page}}, t \rangle$ in (4.5). It is also satisfied for the choices of subsets in figures 5 and 6.
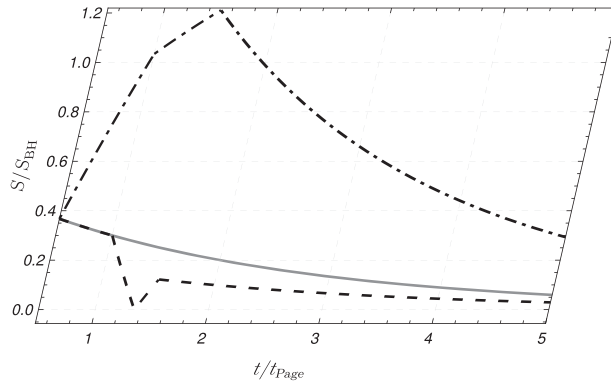
Figure 6. Check of the Araki–Lieb triangle inequality and sub-additivity of entanglement entropy, $S_A + S_B \geqslant S_{AB} \geqslant |S_A - S_B|$, with S_{AB} (continuous), $|S_A - S_B|$ (dashed) and $S_A + S_B$ (dot-dashed). Here we considering, for example, the following two subsets $B = \langle 0, t - \frac{1}{2}t_{\text{Page}} \rangle$ and $A = \langle t, 5t_{\text{Page}} \rangle$.

The final consistency condition we will consider is *strong sub-additivity* which can be stated in terms of non-negativity of the conditional mutual information, or

$$I_{A,BC} \geqslant I_{A,C}. \tag{4.14}$$

This is clearly satisfied in our choice of subsets $\langle 0, t_{\text{Page}} \rangle$ and $\langle t_{\text{Page}}, t \rangle$ because as t increases, the subset B is becomes larger, in other words we can think of this as adding a new subset of the radiation $B \rightarrow BC$, and the mutual information is decreasing. One can check that strong sub-additivity is also satisfied in the example of figure 5 and for other choices of subsets.

To summarize, we do not have general proofs that all the necessary consistency conditions are satisfied by the entropies, but in all checks have tried they are seen to be satisfied.

5. Hunt the purifier

The implication of Page's argument [4] for the behaviour of the entropy of the Hawking radiation is that a late mode B of the radiation must be entangled with an early mode R_B, in order that the initial pure state evolves ultimately back to a pure state when the black hole evaporates away. In this section, we will calculate the mutual information of a small subset of late modes B with subsets of early modes and by maximizing the latter, identify where R_B lies within the early radiation. But then we shall turn the scenario around and ask which future modes B is correlated with (figure 7).

5.1. Searching for R_B in the past

To this end, we pick a subset of Hawking modes B emitted in a small temporal interval $\langle t, t + \delta \rangle$ captured in a spatial interval $[0, \delta]$ at time $t + \delta$. Note by small, we mean that the dimensionless quantity $k\delta$ is small. The question is, what is the smallest subset of the earlier radiation R emitted between times $\langle t_1, t_2 \rangle$, captured in spatial interval $[t + \delta - t_2, t + \delta - t_1]$ at time $t + \delta$, that has maximal mutual information at $\mathcal{O}(k^{-1})$ with the late modes B. This provides a way of identifying where the purifier R_B of the modes B are located in the earlier Hawking radiation
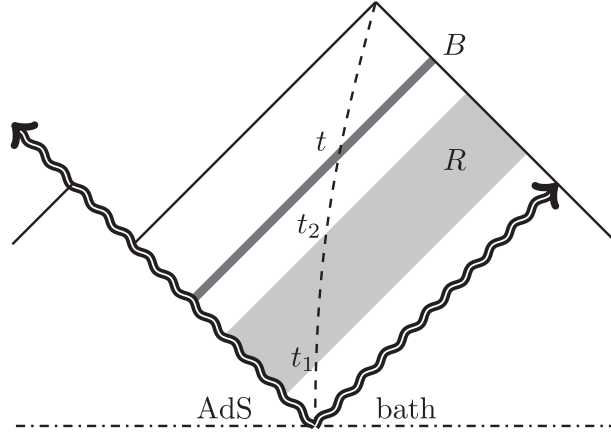
Figure 7. A small set of late Hawking modes B and a subset of the early Hawking radiation R. We are interested in the correlation of B and the smallest region R with the maximum correlation.

The first point to make is that B is a narrow interval, so it is dominated by its no-island saddle, so

$$S_B = k\delta S_{\text{BH}} \; e^{-kt/2}. \tag{5.1}$$

In order for there to be non-vanishing mutual information, R must be dominated by an island saddle. In particular, this requires the island-(1) saddle to dominate and

$$2 \; e^{-kt_1/2} - \frac{3}{2} \; e^{-kt_2/2} > 1. \tag{5.2}$$

However, when R has an island, the combined subsystem BR can be dominated by the island-(3) saddle and the mutual information vanishes because

$$S_{B,\text{no island}} + S_{R,\text{island}(1)} = S_{BR,\text{island}(3)}. \tag{5.3}$$

A non-vanishing $I_{B,R}$ only arises when BR switches over to its island-(1) saddle. The competition between these two determines the mutual information evaluated at time $t + \delta$ (with R at its island-(1) saddle)

$$I_{B,R} = S_{\text{BH}} \; \max\left(0, -3 \; e^{-k(t+\delta)/2} + 4 \; e^{-kt/2} - e^{-kt_2/2}\right). \tag{5.4}$$

Since δ is small, for $I_{B,R}$ to be non-vanishing we require

$$t_2 > t - 3\delta \tag{5.5}$$

and in order that the mutual information is a maximum requires that the early interval is adjacent to the late interval, i.e. $t_2 = t$, giving

$$I_{B,R} = \frac{3}{2}S_B, \tag{5.6}$$

assuming that the cut off term is negligible and δ is small.

Notice that (5.6) is independent of t_1 and so to identify where the entangled partner modes R_B are, we should take t_1 as large as possible compatible with the constraint (5.2) with $t_2 = t$. This determines

$$t_1 = \frac{2}{k} \log \frac{4}{2 + 3 \, \mathrm{e}^{-kt/2}}. \tag{5.7}$$

So the modes that are correlated with the late modes localized in the small interval $B = \langle t, t + \delta \rangle$ lie de-localized in the large interval of earlier modes

$$R_B \subset \left\langle \frac{2}{k} \log \frac{4}{2 + 3 \, \mathrm{e}^{-kt/2}}, t \right\rangle. \tag{5.8}$$

As a consistency check, note that above requires that $t > t_{\mathrm{Page}}$ which is what we would intuitively expect: the black hole must be old. In the next subsection will see that the maximum mutual information is $\frac{3}{2} S_B$ and not $2 S_B$ because B is also entangled with later modes.

5.2. Searching for R_B in the future

We can reverse the logic of the last section, and consider the same scenario but where now R are later Hawking modes. So now we have $t_1 > t$ and the radiation corresponding to R and B are collected in intervals $[0, t_2 - t_1]$ and $[t_2 - t - \delta, t_2 - t]$, respectively, at time t_2. One finds that in order to maximize the mutual information $I_{R,B}$, we need to begin collecting radiation immediately so that $t_1 = t + \delta$.

Then the question is, how small can t_2 be whilst maintaining maximum mutual information $I_{R,B}$. If the modes B are early, i.e. $t < t_{\mathrm{Page}}$, the maximum $I_{R,B}$ is achieved when S_R switches from its no-island to island-(1) saddle. This occurs when (5.2) is satisfied with $t_1 = t$, which means the minimum t_2 is

$$t_2 = \frac{2}{k} \log \frac{3}{4 \, \mathrm{e}^{-kt/2} - 2}. \tag{5.9}$$

Since S_{RB} is also in its island-(1) saddle, we have, for small δ,

$$I_{R,B} = 2 S_B, \tag{5.10}$$

where we are assuming the cut off is negligible.

When $t > t_{\mathrm{Page}}$, the rôles of the island-(1) saddles of R and RB change to the (21) and (41) island saddles, respectively. Now the minimum t_2 is determined by when these saddles dominate over the non-island saddle, i.e. when

$$t_2 = t + \frac{2}{k} \log 3. \tag{5.11}$$

In this case,

$$I_{B,R} = \frac{1}{2} S_B. \tag{5.12}$$

This is a very satisfying result because we know that for $t > t_{\mathrm{Page}}$, B is correlated with the early radiation as in (5.6), i.e. $I_{B,R} = \frac{3}{2} S_B$. So, overall, S_B is correlated with the rest of the Hawking radiation with a mutual information $2 S_B$, which is what one expect if B is entangled with the rest of the radiation and the overall state is pure, at least approximately so when the extremal entropy is negligible.
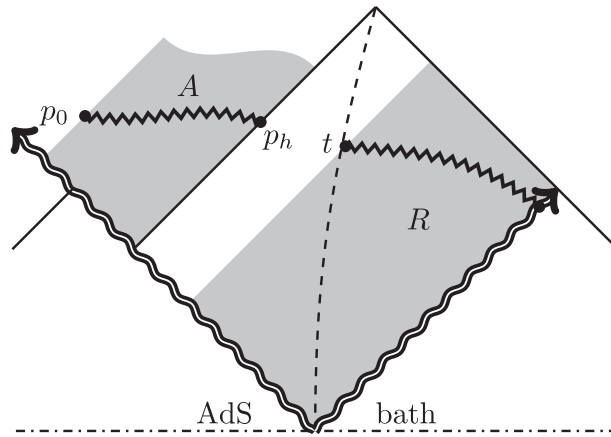
Figure 8. The Hawking radiation R of an old black hole, emitted in the interval $\langle 0, t \rangle$. The modes A behind the horizon are contained in an interval $[p_0, p_{\rm h}]$ where $p_{\rm h}$ is on the horizon.

6. Entanglement-monogamy problem

One way to present the entanglement-monogamy problem is simply the fact that unitarity requires for an old black hole that the Hawking radiation has a falling entropy, in other words, its state is becoming purer, whilst at the same time it must be entangled with modes behind the horizon in order to have a smooth geometry at the horizon. This seems contradictory.

In order to put some flesh on the bones, let us consider the radiation R emitted in the interval $\langle 0, t \rangle$ past the Page time $t > t_{\rm Page}$ and a large subset of the modes behind the horizon A: see figure 8. The argument is similar to that in [13]. Past the Page time, where S_R is decreasing, we can expect—and verify later—that $S_A > S_R$ and so the triangle inequality of Araki and Lieb (1.2) implies

$$S_{AR} \geqslant S_A - S_R. \tag{6.1}$$

This can be written as

$$I_{A,R} \leqslant 2S_R. \tag{6.2}$$

So the fact that the late modes are entangled with the early modes limits the mutual information of the radiation with the modes inside the horizon.

Now in order for the inequality (6.2) to have any teeth, we have to assume that the UV cut is not so large so that $S_{\rm BH} \gg |c \log \varepsilon_{\rm UV}|$ and hence the cut off terms are small compared with the $\mathcal{O}(k^{-1})$ contributions to the entropy. Making this assumption, let us calculate the mutual information $I_{A,R}$. The interval A is the region behind the horizon between points p_0 and $p_{\rm h}$, where p_0 is some point deep inside the black hole and $p_{\rm h}$ is a point on the horizon, i.e. with coordinate $w_{\rm h}^- = 0$. The interval in the bath is $[0, t]$ between points p_1 on the boundary, with coordinate $w^- = -\hat{f}(t)^{-1}$, and p_2 far into the bath.

The first issue to settle is whether the interval behind the horizon A, or AR, can have an island saddle itself. There is no reason, *a priori*, why they cannot. We do not have a general proof that there are no island saddles that can contribute, but suppose we try to have a QES $p_{\hat{1}}$ somewhere near the horizon. The extremization will yield a pair of equations dominated by

the interaction between p_h and the QES[11]:

$$\frac{2\pi}{\beta k} w_{\hat{1}}^+ - \frac{1}{w_{\hat{1}}^- - w_h^-} + \cdots = 0, \qquad \frac{2\pi}{\beta k} w_{\hat{1}}^- - \frac{1}{4 w_{\hat{1}}^+} + \cdots = 0. \tag{6.3}$$

This means that

$$w_{\hat{1}}^+ = -\frac{3\beta k}{8\pi} \cdot \frac{1}{w_h^-}, \qquad w_{\hat{1}}^- = -\frac{3}{4} w_h^- \tag{6.4}$$

and so as p_h approaches the horizon, $w_h^- \to 0$, the QES to pushed off to an unphysical $w_{\hat{1}}^+ \to \infty$, $w_{\hat{1}}^- \to 0$. While this is not a proof that an island cannot dominate, it is suggestive and we will assume that it is true for A and AR.

If R were in its no-island saddle, then the mutual information $I_{A,R}$ would come from the cross terms between the points inside the black hole and those in the bath. However, R *is* in its island-(1) saddle and therefore,

$$I_{A,R} \sim -\frac{c}{6} \log \frac{\sigma_{h1} \sigma_{02}}{\sigma_{01} \sigma_{h2}} + S_{R,\text{island}(1)} - S_{R,\text{no island}}, \tag{6.5}$$

where the σ's are expressed in the (y^+, w^-) frame. We have $S_{R,\text{island}(1)} = S_{\text{BH}} e^{-kt/2}$ and $S_{R,\text{noisland}} = 2 S_{\text{BH}} (1 - e^{-kt/2})$. The cross terms are dominated by the α_{h1} term, and in particular the piece coming from the difference of the w^- coordinates which is becoming small at late times; hence

$$I_{A,R} \sim -\frac{c}{6} \log(\widehat{f}_1^{-1} - 0) - S_{\text{BH}} \left(2 - 3 \, e^{-kt/2} \right)$$

$$\to 4 S_{\text{SBS}}(t) - S_{\text{BH}} \left(2 - 3 \, e^{-kt/2} \right) = S_{\text{BH}} \left(2 - e^{-kt/2} \right). \tag{6.6}$$

This is in contradiction with the expectation (6.2) precisely when $t > t_{\text{Page}}$, and implies some radical departure from the idea that the inside of the black hole is a separate subsystem.

We can put our finger on exactly where the monogamy argument goes wrong. From (6.6), since $S_A = S_{AR} + I_{A,R} - S_R$, this clearly indicates a breakdown of the Araki–Lieb triangle inequality:

$$S_A - S_R = S_{AR} + S_{\text{BH}} \left(2 - 3 \, e^{-kt/2} \right) \not< S_{AR}, \tag{6.7}$$

precisely for $t > t_{\text{Page}}$. Intuitively, what is happening is that S_R is dominated by a saddle that includes the island in the interior of the black hole, but S_{AR} has no island because it already includes the interior region. So the subsystems R and A actually overlap because R has an island, and the triangle inequality only applies when the two subsystems are distinct. Note that the Araki–Lieb inequality was also identified as being at risk in the related scenario in [6].

These considerations lend weight to the so-called '$A = R_B$' hypothesis (e.g. see the review [13]) which identifies the Hilbert space of modes in the early radiation R_B, that are entangled with late modes B, with the Hilbert space of modes behind the horizon A. This leads to the astonishing conclusion that the modes inside an old black hole are actually modes of the Hawking radiation emitted earlier.

---

[11] We ignore the distinction between $\omega_{\hat{1}}^+$ and $w_{\hat{1}}^+$ which is sub-leading in the exponent.

7. Recovering information

In this section, we consider the effect of an additional shockwave inserted on the boundary at a later time. The shockwave has negligible energy, so there is no significant back reaction, but carries entanglement entropy between its out- and in-going components. Since it is information carrying, as is the fashion [5], we will refer to the in-going shockwave as the 'diary' and the out-going one as its 'purifier'.

Before we turn to the gravitational setting, let us consider the effect in a CFT in Minkowski space. Specifically, let us consider the effect on the mutual information of two intervals D and P. As a component of the shockwave, e.g. the diary, enters an interval, say D, the entropies S_D and S_{DP} jump by S_{diary} so the mutual information stays the same. If D contains the diary and P the purifier, both S_D and S_P jump by S_{diary} but S_{DP} stays the same. So the mutual information $I_{D,P}$ jumps by $2S_{\text{diary}}$. This makes intuitive sense, if D and P contain the diary and the purifier, respectively, the mutual information increases by $2S_{\text{diary}}$.

7.1. Black hole as a mirror scenario

Within the JT gravity framework, there are several scenarios we could discuss. The simplest is the one that matches the thought experiments of Hayden and Preskill [5], where a diary, i.e. a local entanglement carrying quench, is dropped into the black hole at some point on boundary at time $t_0 > 0$: see figure 9. All the Hawking radiation is collected up to some time t and the question is how large does t have to be to recover the information of diary. To this end, we reserve a narrow interval in bath $P = [t - t_0 - \delta, t - t_0 + \delta]$, designed to pick up the purifier at time t, where δ is small. The complement $D = [0, t - t_0 - \delta] \cup [t - t_0 + \delta, t]$, is then the Hawking radiation collected in this context. The information has returned at time t if

$$I_{D,P}(t) \approx 2S_{\text{diary}}. \tag{7.1}$$

In order for the information to be returned, requires that the islands of D and DP contain the diary, whereas P must not have an island[12]. This latter requirement is guaranteed since P is a narrow interval and, as long as the entropy of the diary is not too large compared with S_{BH}, the no-island saddle always dominates:

$$
\begin{aligned}
S_{P,\text{no island}} &= 2S_{\text{BH}} \, e^{-kt_0/2}(e^{k\delta/2} - e^{-k\delta/2}) + S_{\text{diary}} \approx S_{\text{diary}}, \\
S_{P,\text{island}(1)} &= S_{\text{BH}} \left(2 - e^{-kt_0/2}(2 \, e^{k\delta/2} - e^{-k\delta/2})\right) \approx S_{\text{BH}}(2 - e^{-kt_0/2}),
\end{aligned} \tag{7.2}
$$

In the above, note that the no-island saddle is shifted by S_{diary} because P contains the purifier but the island contains the diary and so the island saddle is not shifted. We are assuming that diary entropy is not too large compared with the black hole so $\eta = S_{\text{diary}}/S_{\text{BH}} \ll 1$.

So in order for the information to be recovered, D and DP must be dominated by their island-(1) saddles and the diary must be contained within the islands, so that

$$S_{DP} \approx S_D - S_{\text{diary}}, \tag{7.3}$$

from which (7.1) follows.

---

[12] There is a subtlety here because D consists of two intervals, albeit ones separated by a small amount, and so has two island saddles: in the notation of (3.31) island-(1) and island-(3). In this case, the saddle (3) is always sub-dominant. Then the non-diary parts of S_D and S_{DP} for the no-island and island saddles are approximately the same.
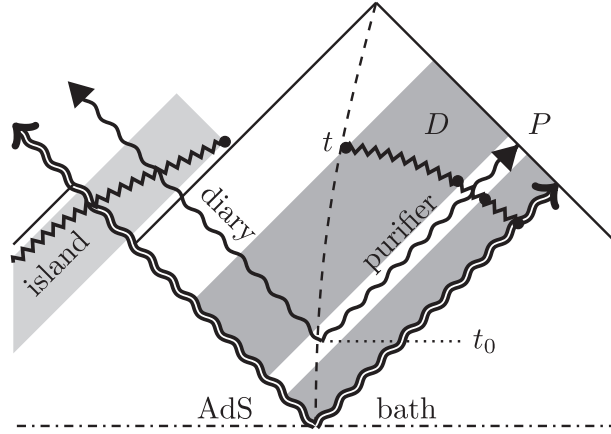
Figure 9. In this scenario, the Hawking radiation is collected in interval *D* from the beginning of the evaporation, excluding a small interval *P* that is used to collect the purifier of the diary. Also shown is the island of *D* that dominates when *t* is large enough. The information is returned in *D* when *t* is a little larger than the Page time determined by the diary's entropy.

Let us consider *D*. Although this consists of two intervals, the non-diary part of the entropy is approximately that of a single interval $[0, t]$, since δ is small. Of course *DP* is the whole interval $[0, t]$. The difference is that when the diary is in the island of *D* the island entropy is shifted *up* by S_{diary}, whereas for *DP* the opposite is true, the no-island entropy is shifted by S_{diary} since it contains the purifier. So the island for *DP* appears at t_-, *before* that of *D* at time t_+, where

$$2 - 2\, e^{-kt_{\mp}/2} = e^{-kt_{\mp}/2} \mp \eta \quad \implies \quad t_{\mp} \approx t_{\text{page}} \mp \frac{\eta}{k}. \tag{7.4}$$

However, we have to check that the diary is in the island, that is the w^+ coordinate of the diary is less than that of the QES p_2 of *DP*, or *D*, at time *t* when the radiation is collected, i.e.

$$\hat{f}(t_0) < \frac{3\beta k}{8\pi}\hat{f}(t)e^{kt/2}. \tag{7.5}$$

The pre-factor here, leads to a delay $t > t_0 + \Delta t_{\text{scr.}}$,

$$\Delta t_{\text{scr.}} \approx \frac{\beta}{2\pi}e^{kt/2}\ \log\left(\frac{8\pi}{3\beta k}e^{-kt/2}\right), \tag{7.6}$$

identified as the scrambling time of the black hole, however, this is a small effect when the leading time scales are of order k^{-1} and so, to leading order, we ignore it. Collecting all this together yields the mutual information, which increases in a piece-wise continuous fashion as

$$I_{D,P}(t) = \begin{cases} 0 & t \leqslant t_-, \\ S_{\text{BH}}\left(2 - 3\,e^{-kt/2}\right) + S_{\text{diary}} & t_- \leqslant t \leqslant t_+, \\ 2S_{\text{diary}} & t_+ \leqslant t. \end{cases} \tag{7.7}$$

So the following picture emerges that is entirely consistent with the analysis of Hayden and Preskill [5]. If we collect all the Hawking radiation emitted from $t = 0$, then if the diary is thrown in before the time t_-, a little earlier than the Page time, then the information will be returned a little later than the Page time at t_+. From (7.4), note that the additional time beyond t_Page is proportional to the diary's entropy. If the diary is thrown in after t_+, then it is returned almost immediately after a short delay given by the instantaneous scrambling time of the black hole $\Delta t_\text{scr.}$ at time t_0. This is the origin of the 'black hole as a mirror' metaphor.

7.2. Recovery after the fact

There are variations of this scenario we can analyse. Suppose we throw the diary in, but only then start to collect the Hawking radiation in a temporal interval $\langle t_0, t \rangle$, as shown in figure 10. In this case, we collect the Hawking radiation in a spatial interval $D = [0, t - t_0]$ at time t. Now the conditions (7.4) that DP and D have an island at $t = t_\mp$, respectively, are modified to

$$2\,e^{-kt_0/2} - 2\,e^{-kt_\mp/2} = e^{-kt_\mp/2} + 2 - 2\,e^{-kt_0/2} \mp \eta$$
$$\implies \quad t_\mp = t_0 + \frac{2}{k} \log \frac{3}{4 - (2 \mp \eta)e^{kt_0/2}}. \tag{7.8}$$

We know that in order to recover the information at time t, i.e. $I_{D,P}(t) \approx 2S_\text{diary}$, requires $t \geqslant t_+$. It follows that the diary cannot be thrown in too late

$$t_0 < \frac{2}{k} \log \frac{4}{2 + \eta}. \tag{7.9}$$

In addition, the minimum collection time $t_+ - t_0$ becomes longer for later t_0:

$$t_+ - t_0 = \frac{2}{k} \log \frac{3}{4 - (2 + \eta)e^{kt_0/2}}. \tag{7.10}$$

Note that if the diary is thrown in close to, or later than, the bound (7.9), t_+ becomes too large or diverges and our approximations will break down. One expects that the exact analysis will show that the information does come out eventually, as the black hole evaporates back to the extremal black hole.

8. Behind the horizon in the bath

The entropy calculations show that for an old black hole the physics behind the horizon is subtly encoded in the state of the bath. This extends to more refined quantities like the entropy of an interval in the bath in an excited state of the CFT created by the local operator insertion, as our story of recovering information in section 7 shows. In this section, we consider this in a little more detail. The question we ask, is how does the entropy of an interval of radiation in the bath respond to an operator insertion behind the horizon?

First of all, in the non-gravitational setting, the response of an interval to an operator insertion has been well studied, e.g. [45, 46]. The excited state is $\mathcal{O}(y_0^\pm + i\varepsilon)\,|0\rangle$, defined as in (2.3) with a small imaginary shift in the insertion point in order to ensure that the state is normalized. If the interval is $A = [b_1, b_2]$ and the operator is inserted at point $y_0^\pm$, then the entropy $\Delta S_A(t)$, the difference of the entropy with and in the absence of the operator insertion, responds in a causal way. This means that $\Delta S_A(t)$ becomes non-trivial approximately when either $t - b_2 \leqslant y_0^- \leqslant t - b_1$ or $t + b_1 \leqslant y_0^+ \leqslant t + b_2$.
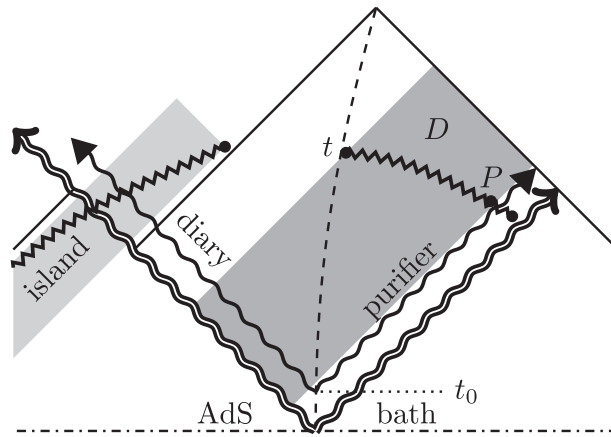
Figure 10. In this scenario, the Hawking radiation is collected after the diary is thrown into the black hole at t_0. The information is returned in D if t_0 is not too large and t is sufficiently late.

The intuitive picture is that the operator insertion creates a left- and a right-moving shockwave and the entropy of the interval responds as either of these shockwaves moves through the interval. The leading order effect for small ε, at $\mathcal{O}(\varepsilon^0)$, is that $\Delta S_A(t)$ jumps by log $d_{\mathcal{O}}$, where $d_{\mathcal{O}}$ is the *quantum dimension* of the operator—not to be confused with its scaling dimension—when either of the shockwave moves into the interval. The quantity log $d_{\mathcal{O}}$ is a measure of the entanglement of the operator $\mathcal{O}$ between the left- and right-moving shockwaves. In section 7, we called log $d_{\mathcal{O}}$ the entropy of the diary. There are interesting universal corrections in an expansion in the small quantity ε [46].

Given how the entropy of a subregion behaves in the presence of an operator insertion, it seems obvious that, when an interval in the bath is in an island saddle, operator insertions behind the horizon that pass through the island will be witnessed by the interval in the bath. This seems to be a prima facie violation of locality that is, in principle, measurable in the bath in the sense that we describe in the discussion section 1.1.

Let us consider the effect in more detail. Let us take an interval in the bath $A = [b_1, b_2]$. At late times, and certainly $t > b_2$, the entropy of the interval involves a competition between the no-island and the island-(21) saddles:

$$S_{\text{no island}} = 2S_{\text{BH}} \, e^{-kt/2} \left(e^{kb_2/2} - e^{kb_1/2} \right),$$

$$S_{\text{island(21)}} = S_{\text{BH}} \, e^{-kt/2} \left(e^{kb_1/2} + e^{kb_2/2} \right). \tag{8.1}$$

The island-(21) saddle dominates when $e^{kb_2/2} > 3 \, e^{kb_1/2}$. This saddle illustrated in figures 2 and 11.

Suppose the operator insertion behind the horizon as shown in figure 11, at $w_0^{\pm}$. In this case, the right-moving shockwave lies in the island if

$$\frac{1}{3\hat{f}(t - b_2)} \leqslant w_0^- \leqslant \frac{1}{3\hat{f}(t - b_1)}. \tag{8.2}$$

So the effect is indistinguishable to a scenario where the operator insertion in made in the bath at a point with coordinate $y_0^- = \hat{f}^{-1}(1/(3w_0^-))$.

26
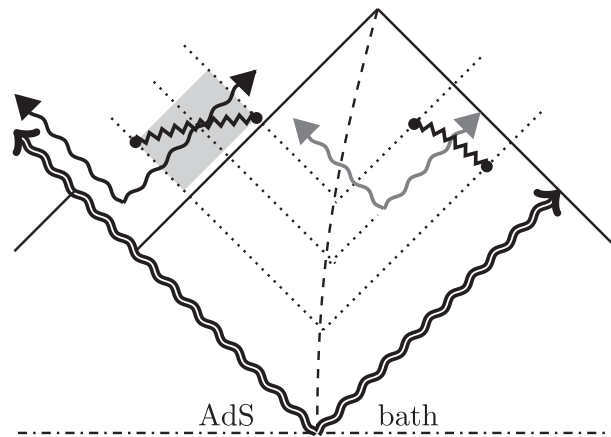
Figure 11. An operator insertion made behind the horizon generates a pair of shock-waves. Here, at a certain time, the right moving one is in the island of the interval in the bath, the shaded region, and creates an effect that is indistinguishable from an operator insertion made in the bath as shown.

9. Discussion

In this work we set up the calculation of von Neumann entropies for an arbitrary number of subsystems within the Hawking radiation bath of an evaporating black hole in JT gravity produced by a shockwave injected via a local quench. The primary goal in this analysis was to display the correlations within early and late Hawking modes, and to flesh out the '$A = R_B$' scenario within this framework. The slow evaporation limit $k \ll 1$ with kt fixed, in conjunction with large interval sizes $\sim \mathcal{O}(k^{-1})$ results in a simple but rich, analytically tractable interplay of island and no-island saddles. We focussed attention on island saddles with only two QESs since we generically expect additional QES contributions to significantly increase the entropy. Such islands are sufficient for teasing out the refined correlations between two disjoint subsets of the Hawking modes in the bath. Whether islands with additional QESs dominate, likely depends on the number and size of separations between the intervals in the bath, and it would be very interesting to understand the conditions under which such saddles could become relevant, and how multi-QES islands may compete with each other. In the bath CFT, Rényi/entanglement entropies can be understood as correlators of twist fields. The operator product of twist fields is dominated by the stress tensor in the short interval limit. It would be interesting to understand what implications the calculation of multi-interval entropies has for correlators of local operators (such as the stress tensor) and lack of cluster property (see e.g. [47]) in the presence of replica wormhole (island saddle) contributions.

Acknowledgments

ORCID iDs

Timothy J Hollowood ⬤ https://orcid.org/0000-0002-3258-320X

References

[1] Penington G, Shenker S H, Stanford D and Yang Z 2019 Replica wormholes and the black hole interior (arXiv:1911.11977 [hep-th])

[2] Almheiri A, Hartman T, Maldacena J, Shaghoulian E and Tajdini A 2020 Replica wormholes and the entropy of Hawking radiation *J. High Energy Phys.* JHEP05(2020)013

[3] Hawking S W 1976 Breakdown of predictability in gravitational collapse *Phys. Rev.* D **14** 2460

[4] Page D N 1993 Information in black hole radiation *Phys. Rev. Lett.* **71** 3743

[5] Hayden P and Preskill J 2007 Black holes as mirrors: quantum information in random subsystems *J. High Energy Phys.* JHEP09(2007)120

[6] Almheiri A, Engelhardt N, Marolf D and Maxfield H 2019 The entropy of bulk quantum fields and the entanglement wedge of an evaporating black hole *J. High Energy Phys.* JHEP12(2019)063

[7] Penington G 2019 Entanglement wedge reconstruction and the information paradox (arXiv:1905.08255 [hep-th])

[8] Almheiri A, Mahajan R and Maldacena J 2019 Islands outside the horizon (arXiv:1910.11077 [hep-th])

[9] Gautason F F, Schneiderbauer L, Sybesma W and Thorlacius L 2020 Page curve for an evaporating black hole *J. High Energy Phys.* JHEP05(2020)091

[10] Hashimoto K, Iizuka N and Matsuo Y 2020 Islands in Schwarzschild black holes *J. High Energy Phys.* JHEP06(2020)085

[11] Hollowood T J and Kumar S P 2020 Islands and page curves for evaporating black holes in JT gravity (arXiv:2004.14944 [hep-th])

[12] Almheiri A, Marolf D, Polchinski J and Sully J 2013 Black holes: complementarity or firewalls? *J. High Energy Phys.* JHEP02(2013)062

[13] Harlow D 2016 Jerusalem lectures on black holes and quantum information *Rev. Mod. Phys.* **88** 015002

[14] Harlow D and Hayden P 2013 Quantum computation vs firewalls *J. High Energy Phys.* JHEP06(2013)085

[15] Susskind L and Thorlacius L 1994 Gedanken experiments involving black holes *Phys. Rev.* D **49** 966–74

[16] Bousso R 2013 Complementarity is not enough *Phys. Rev.* D **87** 124023

[17] Banks T and Fischler W 2012 Holographic space–time does not predict firewalls (arXiv:1208.4757 [hep-th])

[18] Papadodimas K and Raju S 2013 An infalling observer in AdS/CFT *J. High Energy Phys.* JHEP10(2013)212

[19] Verlinde E and Verlinde H 2013 Black hole entanglement and quantum error correction *J. High Energy Phys.* JHEP10(2013)107

[20] Maldacena J and Susskind L 2013 Cool horizons for entangled black holes *Fortschr. Phys.* **61** 781–811

[21] Papadodimas K and Raju S 2014 Black hole interior in the holographic correspondence and the information paradox *Phys. Rev. Lett.* **112** 051301

[22] Papadodimas K and Raju S 2014 State-dependent bulk-boundary maps and black hole complementarity *Phys. Rev.* D **89** 086010

[23] Verlinde E and Verlinde H 2013 Behind the horizon in AdS/CFT (arXiv:1311.1137 [hep-th])

[24] Araki H and Lieb E H 1970 Entropy inequalities *Commun. Math. Phys.* **18** 160–70

[25] Engelsöy J, Mertens T G and Verlinde H 2016 An investigation of AdS$_2$ backreaction and holography *J. High Energy Phys.* JHEP07(2016)139

[26] van Enk S J and Beenakker C W J 2012 Measuring Tr ρ^n on single copies of ρ using random measurements *Phys. Rev. Lett.* **108** 110503

[27] Elben A, Vermersch B, Roos C F and Zoller P 2019 Statistical correlations between locally randomized measurements: a toolbox for probing entanglement in many-body quantum states *Phys. Rev.* A **99** 052323

[28] Jackiw R 1985 Lower dimensional gravity *Nucl. Phys.* B **252** 343

[29] Teitelboim C 1983 Gravitation and Hamiltonian structure in two spacetime dimensions *Phys. Lett.* B **126** 41

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

[31] Casini H and Huerta M 2009 Reduced density matrix and internal dynamics for multicomponent regions *Class. Quantum Grav.* **26** 185005
[32] Calabrese P and Cardy J 2004 Entanglement entropy and quantum field theory *J. Stat. Mech.* P06002
[33] Calabrese P and Cardy J 2005 Evolution of entanglement entropy in one-dimensional systems *J. Stat. Mech.* P04010
[34] Ryu S and Takayanagi T 2006 Holographic derivation of entanglement entropy from AdS/CFT *Phys. Rev. Lett.* **96** 181602
[35] Hubeny V E, Rangamani M and Takayanagi T 2007 A Covariant holographic entanglement entropy proposal *J. High Energy Phys.* JHEP07(2007)062
[36] Faulkner T, Lewkowycz A and Maldacena J 2013 Quantum corrections to holographic entanglement entropy *J. High Energy Phys.* JHEP11(2013)074
[37] Engelhardt N and Wall A C 2015 Quantum extremal surfaces: holographic entanglement entropy beyond the classical regime *J. High Energy Phys.* JHEP01(2015)073
[38] Cerf N J and Adami C 1997 Negative entropy and information in quantum mechanics *Phys. Rev. Lett.* **79** 5194
[39] Tucci R R 2002 Entanglement of distillation and conditional mutual information (arXiv:quant-ph/0202144)
[40] Christandl M and Winter A 2004 'Squashed entanglement': an additive entanglement measure *J. Math. Phys.* **45** 829
[41] Carlen E A and Lieb E H 2012 Bounds for entanglement via an extension of strong subadditivity of entropy *Lett. Math. Phys.* **101** 1
[42] Hayden P, Headrick M and Maloney A 2013 Holographic mutual information is monogamous *Phys. Rev.* D **87** 046003
[43] Li T, Chu J and Zhou Y 2020 Reflected entropy for an evaporating black hole (arXiv:2006.10846 [hep-th])
[44] Takayanagi T and Umemoto K 2018 Entanglement of purification through holographic duality *Nat. Phys.* **14** 573
[45] He S, Numasawa T, Takayanagi T and Watanabe K 2014 Quantum dimension as entanglement entropy in two dimensional conformal field theories *Phys. Rev.* D **90** 041701
[46] David J R, Khetrapal S and Kumar S P 2016 Universal corrections to entanglement entropy of local quantum quenches *J. High Energy Phys.* JHEP08(2016)127
[47] Blommaert A 2020 Dissecting the ensemble in JT gravity (arXiv:2006.13971 [hep-th])