Islands in asymptotically flat 2D gravity

Thomas Hartman,[a] **Edgar Shaghoulian**[a] **and Andrew Strominger**[b]

[a]*Department of Physics, Cornell University,*
Ithaca, New York, U.S.A.

[b]*Department of Physics, Harvard University,*
Cambridge, MA, U.S.A.

E-mail: hartman@cornell.edu, eshaghoulian@cornell.edu,
strominger@physics.harvard.edu

ABSTRACT: The large-N limit of asymptotically flat two-dimensional dilaton gravity coupled to N free matter fields provides a useful toy model for semiclassical black holes and the information paradox. Analyses of the asymptotic information flux as given by the entanglement entropy show that it follows the Hawking curve, indicating that information is destroyed in these models. Recently, motivated by developments in AdS/CFT, a semiclassical island rule for entropy has been proposed. We define and compute the island rule entropy for black hole formation and evaporation in the large-N RST model of dilaton gravity and show that, in contrast, it follows the unitary Page curve. The relation of these two observations, and interesting properties of the dilaton gravity island rule, are discussed.

KEYWORDS: 2D Gravity, Black Holes, Models of Quantum Gravity

ARXIV EPRINT: 2004.13857

Contents

1 Introduction

Many years ago, Gibbons and Hawking [1] showed that the entropy of a black hole can be formally but simply calculated in the semiclassical limit from the Euclidean path integral. That calculation is now understood, at least in some cases, to correctly enumerate the quantum microstate degeneracy of the black hole. It is a wonderful and deep surprise that semiclassical gravity is smart enough to reproduce this degeneracy. In the intervening decades, the unreasonable efficacy of semiclassical gravity has been repeatedly demonstrated in disparate contexts. Most cases however involve symmetries of one kind or another — such as time translations. Recently [2–6] a remarkable proposal has been made invoking semiclassical gravity to analyze the information flow out of black holes. The starting point is the famous and well-understood Ryu-Takayanagi formula [7] for anti-deSitter space (AdS), which semiclassically computes microscopic quantum entanglement entropies in terms of areas of extremal surfaces. Recent proposed generalizations of this formula define an 'island rule entropy' for time-dependent evaporating black holes in AdS

— and even for those in flat space — involving a 'quantum extremal surface' or QES [8–10]. The island rule entropy elegantly reproduces at leading order the Page curve describing unitary information flux, accompanied by a refined semiclassical picture of the evaporation process. We hope that this significant progress will provide important clues to guide us to a complete microscopic resolution of the information paradox.

In a separate direction, the process of black hole evaporation has been systematically studied starting in the early 90s in the context of the large-N limit of 2D dilaton gravity minimally coupled to N free matter fields. A review can be found in [11]. The first such model, known as CGHS [12], originated as a dimensional truncation of the near extremal NS fivebrane in string theory, but the CGHS model and its variants [13–16] can also be studied as toy 2D black hole models in their own right. These models are of interest because the large-N limit includes Hawking radiation as well as the backreaction on the metric, and yet is soluble. The quantum state on $\mathcal{I}^+$, the information flux, or equivalently entanglement entropy as a function of retarded time can be numerically or in some models (in particular the RST model [14]) analytically computed. The information flux at $\mathcal{I}^+$ follows the Hawking curve, not the Page curve. The models therefore seem to destroy information.

In assessing this uncomfortable conclusion it is important to realize that the dilaton gravity models, despite some efforts, were never derived either as a decoupling limit of string theory or as the large-N limit of an *exactly* defined model of 2D gravity. In stringy limits, typically an infinite tower of interacting fields do not decouple. Since 2D quantum gravity is renormalizable (in fact equivalent to 2D CFT) one might hope to define and study such models in their own right without recourse to a decoupling limit. The 2D bulk theory is in some cases fully soluble at the quantum level (as a null variant of Liouville theory [13, 16]). However, a faithful toy model for black holes requires the dilaton field to be positive, and a boundary is needed to eliminate regions where it becomes negative. Efforts to implement such a boundary condition and avoid naked singularities in a self-consistent manner never fully succeeded. The apparent destruction of information, together with the absence of any known exact definition at finite N, suggest that the large-N dilaton gravity models are not the large N limit of any exact fully consistent 2D quantum theory of gravity.[1] However the jury is still out concerning the best perspective on these models and their relevance to the information paradox.

It would certainly be of great interest to find *some* fully consistent 2D toy model to study black hole formation/evaporation.[2] It remains possible that some modification of CGHS — perhaps involving wormholes, boundary modes or other additional degrees of freedom — exists as an exact quantum theory and obeys the island rule descibed herein.

Of special interest in this regard are the near extremal NS fivebranes, which contain CGHS plus additional massive fields which survive the scaling limit and are holographically

---

[1]A further reason for doubting the existence of such theories is the exact global flavor symmetry, which is not expected in quantum gravity. It is hard to see how the information about the flavor of the matter which collapses to a black hole could ever be retrieved from the Hawking radiation at $\mathcal{I}^+$.

[2]Black holes probably cannot be formed in the $c = 1$ matrix model [17], and the bulk description of the SYK model [18, 19] involves a tower of low-mass states.

dual to little string theory [12, 20, 21]. While these theories are not fully two-dimensional and do not have N matter fields, they can be semiclassically studied at large black hole mass M. As they do have a holographic dual, it is plausible that the information flux at $\mathcal{I}^+$ follows the Page curve, and the island rule entropy as described in this paper agrees with the information flux measured by an asymptotic observer. It would be very interesting to derive this explicitly.

In this paper we compare in detail in 2D these old and new semiclassical analyses of black hole evaporation. We analyze the QES and associated semiclassical island rule entropy in the RST model, employing a definition of the island rule entropy which closely mimics the ones used in AdS [5] but lives at $\mathcal{I}^+$. The explicit expression as a function of retarded time is shown to reproduce the leading-order Page curve, which is a stringent consistency check of the island rule. This of course disagrees with the actual measurable information flux of the RST model which follows the Hawking curve. We analyze the difference and try to learn something from it.

More generally, the semiclassical island rule does not agree with the semiclassical Hawking calculation for 4D black holes even though they apparently have the same regime of validity. If the former is correct, one would like to find the error in the latter. Our hope is that the 2D context considered here proves a useful one for resolving this tension.

The island rule supplies a surprising canonical map from a time along the QES curve in the black hole interior to a retarded time on $\mathcal{I}^+$. It implicitly asserts that when an infalling object crosses the horizon and reaches the QES curve, information about its quantum state becomes available on $\mathcal{I}^+$. This assertion reproduces the Hayden-Preskill scrambling time. The QES is seen at $\mathcal{I}^+$ to hit the singularity a time of order one before the evaporation endpoint. The island rule entropy then has no classical gravitational piece, and reduces to the ordinary quantum entanglement of the surface extending from the QES endpoint to $\mathcal{I}^+$. This surface magically has zero net quantum entanglement to leading order in M due to a cancellation of the order M bulk term with corrections due to the UV cutoff redshift at its boundaries. This provides another self-consistency check of the island rule.

This paper is organized as follows. Section 2 is a lightning review of the needed features of 2D asymptotically flat dilaton gravity and the RST model. The linear dilaton, eternal black hole and evaporating black hole solutions are presented, followed by the microscopic entanglement entropy and the information flux at $\mathcal{I}^+$ for the evaporating black hole. Section 3 begins with the definition of the island rule entropy, which adds to the asymptotic entanglement entropy an island region which ends at the QES with an area-law type gravitational contribution. This formula is then applied to the evaporating black hole and shown to have several remarkable properties, including reproducing the Page curve and the scrambling time. Section 4 considers the eternal black hole, which has a mathematically simplified, if less physically sharp, version of the information paradox: the entanglement entropy of Hawking quanta outside the black hole grows without bound and eventually exceeds the Bekenstein-Hawking black hole entropy. This paradox is beautifully resolved by assuming the island rule. In the appendix we consider the gravitational replica calculation and show that the local equations of motion fix the location of the QES, although we will not construct the wormhole saddles globally.

Other recent work addressing islands and information recovery in black holes includes [22–40]. As this work was nearing completion two papers appeared with overlapping results [41, 42].

2 Review of the RST model

This section contains a lightning review of relevant features of the RST model [14], largely following the conventions of [43] to which we refer the reader for further details. For earlier results on black holes in two dimensions see [44, 45].

2.1 Large-N action and equation of motion

The classical action of the RST model is[3]

$$
S_{cl} = \frac{1}{2\pi} \int d^2x \sqrt{-g} \left[e^{-2\phi}(R + 4(\nabla\phi)^2 + 4\lambda^2) - \frac{N}{24}\phi R \right] + S_{\text{CFT}}
$$
$$
S_{\text{CFT}} = -\sum_{k=1}^{N} \frac{1}{4\pi} \int d^2x \sqrt{-g}(\nabla f_k)^2. \tag{2.1}
$$

To avoid equation clutter we henceforth choose units in which

$$
\lambda = 1. \tag{2.2}
$$

In the quantum effective action, the conformal anomaly contributes a term

$$
S_{\text{anom}} = -\frac{N}{96\pi} \int d^2x \sqrt{-g} R \Box^{-1} R. \tag{2.3}
$$

The large-N limit is $N \to \infty$ with $Ne^{2\phi}$ held fixed. It described by the sum

$$
S_{\text{eff}} = S_{\text{cl}} + S_{\text{anom}} + S_{\text{CFT}}. \tag{2.4}
$$

This action simplifies in conformal gauge $ds^2 = -e^{2\rho}dx^+dx^-$, with $x^\pm = x^0 \pm x^1$ and $R_{+-} = -2\partial_+\partial_-\rho$. In these coordinates the large-N effective action becomes

$$
S_{\text{eff}} = \frac{1}{2\pi} \int d^2x \left[e^{-2\phi} \left(4\partial_+\partial_-\rho - 8\partial_+\phi\partial_-\phi + 2e^{2\rho} \right) + \frac{N}{6}(\rho - \phi)\partial_+\partial_-\rho \right] + S_{\text{CFT}},
$$
$$
S_{\text{CFT}} = \frac{1}{2\pi} \sum_{k=1}^{N} \int d^2x \, \partial_+ f_k \partial_- f_k \tag{2.5}
$$

It is convenient to define

$$
\Omega = \frac{12}{N}e^{-2\phi} + \frac{\phi}{2} - \frac{1}{4}\ln\frac{48}{N}, \qquad \chi = \frac{12}{N}e^{-2\phi} + \rho - \frac{\phi}{2} + \frac{1}{4}\ln\frac{3}{N}. \tag{2.6}
$$

---

[3]More generally, replacing S_{CFT} with any family of 2D CFTs with central charge $c = N$ minimally coupled to gravity would lead to few changes in the following discussion. The ϕR term in (2.1) distinguishes the RST from the CGHS model [12] and was added to restore the conserved current $\partial_\mu(\rho - \phi)$ broken by the anomaly. The existence of this conserved current conveniently simplifies many formulae but does not affect the basic structure, see e.g. [46]. Indeed for our primary focus of the large mass limit there is no difference.

The resulting action

$$S_{\text{eff}} = \frac{N}{12\pi} \int d^2x \left(\partial_+\Omega\partial_-\Omega - \partial_+\chi\partial_-\chi + e^{2\chi-2\Omega} \right) + S_{\text{CFT}} \tag{2.7}$$

then scales with an overall factor of N in our large-N limit where Ω and χ are held fixed.

Note that for real ϕ, Ω obeys

$$\Omega \geq \frac{1}{4}. \tag{2.8}$$

In order to identify S_{eff} as a theory of gravity with black holes, it is essential that the restriction of real ϕ be enforced so that the coefficient of the Einstein action is positive. When 2D dilaton gravity is constructed as a spherical dimensional reduction of higher-dimensional gravity, Ω is the (quantum) area of the spheres, and the condition (2.8) marks the origin of the spacetime where the spheres shrink to zero. If we don't make such a restriction, (2.7) describes a soluble, null variant of Liouville theory [13, 16], but it has no singularities and no black holes.[4] Furthermore, without this restriction the collapsing matter configurations described below will radiate forever to negative infinite mass, whereas with the restriction we can motivate a pasting of the linear dilaton vacuum at the evaporation endpoint, i.e. the point where the $\Omega = 1/4$ curve turns from spacelike to timelike. This cuts off the evaporation. In order to have a theory with black holes, we therefore impose (2.8) as in [14, 43]. This condition looks sensible at large N, but certainly spoils the full quantum solubility.[5] We further impose reflecting boundary conditions on the N quantum matter fields f_k which descends from their reflection through the origin in higher dimensions.

The S_{eff} equations of motion for χ and Ω are $\partial_+\partial_-\Omega = \partial_+\partial_-\chi = -e^{2\chi-2\Omega}$, so $\chi = \Omega + \alpha^+(x^+) + \alpha^-(x^-)$. The residual on-shell conformal diffeomorphism symmetry of conformal gauge allows us to choose coordinates in which

$$\chi = \Omega. \tag{2.9}$$

For reasons which will become apparent we refer to these as Kruskal coordinates and will denote the non-scalar quantities χ and ρ in Kruskal gauge as χ_K and ρ_K. (2.9) implies

$$\phi = \rho_K - \frac{1}{2}\ln\frac{N}{12}$$
$$\Omega = e^{-2\rho_K} + \frac{1}{2}(\rho_K - \ln 2). \tag{2.10}$$

The equation of motion in this gauge is simply

$$\partial_+\partial_-\Omega = -1. \tag{2.11}$$

There are also constraint equations, which are given by varying (2.1) with respect to $g_{\pm\pm}$ before fixing to conformal gauge:

$$\partial_\pm^2\Omega = -T_{\pm\pm}^f - t_\pm, \tag{2.12}$$

---

[4]Some versions of JT gravity have this feature and hence may need supplemental restrictions to provide good models for black hole physics.

[5]Whether any such constraint makes sense at finite N is an open question, studied in [47, 48].
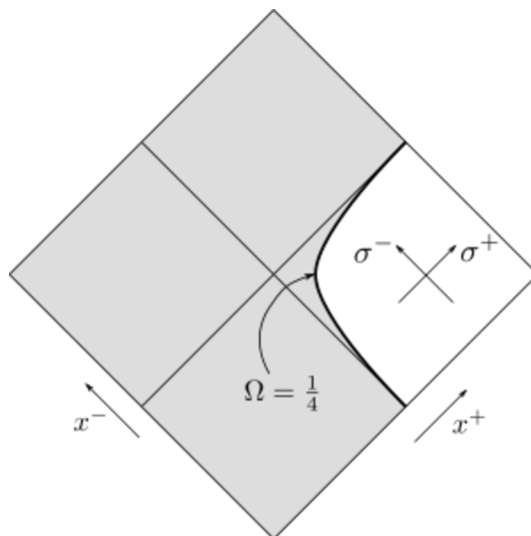
Figure 1. Linear dilaton vacuum. The physical spacetime is the unshaded region, with a boundary at $\Omega = \frac{1}{4}$.

where

$$T^f_{\pm\pm} = \frac{6}{N} \sum_{k=1}^{N} \int d^2x \, \partial_{\pm} f_k \partial_{\pm} f_k \tag{2.13}$$

is the matter stress tensor rescaled by a factor of $\frac{12\pi}{N}$ for notational convenience. $t_\pm$ depends on the choice of coordinates prescribed for normal ordering. If we choose $x'^+(x^+)$ instead of x^+, t_+ shifts by the Schwarzian (see e.g. [43])

$$t'_+ = (\partial'_+ x^+)^2 t_+ + \sqrt{\partial'_+ x} \partial'^2_+ \sqrt{\partial_+ x'^+}, \tag{2.14}$$

with a similar relation for t_-. In the black hole collapse studied below the ambiguity will be uniquely fixed by demanding the absence of incoming particles in inertial frames in the far past.

2.2 Solutions

The general solution to (2.11)–(2.12) can be obtained by integration, although we will only need the following special cases.

2.2.1 Linear dilaton vacuum

The vacuum of the theory, known as the linear dilaton vacuum, in Kruskal coordinates (2.9) is given by

$$\Omega = -x^+ x^- - \frac{1}{4}\ln(-4x^+ x^-) \tag{2.15}$$

with the classical matter stress tensor $T^f_{\pm\pm} = 0$. The physical region of the spacetime with $\Omega \geq \frac{1}{4}$ is covered by the coordinate region with $x^+ x^- \leq \frac{1}{4}$ and $\pm x^\pm > 0$, as illustrated in figure 1. Using (2.9), this implies the 2D Minkowski line element

$$ds^2 = \frac{dx^+ dx^-}{x^+ x^-}. \tag{2.16}$$

From the fact that $\partial_\pm^2 \Omega = \frac{1}{4(x^\pm)^2}$, we conclude that

$$t_\pm = -\frac{1}{4(x^\pm)^2}. \tag{2.17}$$

We will also employ "Minkowski coordinates" defined by

$$x^\pm = \pm e^{\pm\sigma^\pm} \tag{2.18}$$

in which the line element takes the standard flat form

$$ds^2 = -d\sigma^+ d\sigma^-, \tag{2.19}$$

the dilaton is linear

$$\phi = \frac{1}{2}\left(\sigma^- - \sigma^+ - \ln\frac{N}{12}\right), \tag{2.20}$$

and $\Omega = \chi + \frac{1}{2}(\sigma^- - \sigma^+)$. The standard Minkowski vacuum is annihilated by Minkowski modes which are negative frequency with respect to Minkowski time $\sigma^+ + \sigma^-$. Using the transformation law (2.14) we find that $t_\pm = 0$ in the Minkowski frame and (2.15) hence corresponds to the quantum fields f_k in the standard Minkowski vacuum.

2.2.2 Eternal black hole

The solution for the eternal black hole (see figure 7) is

$$\Omega = -x^+ x^- + M, \tag{2.21}$$

with $T^f_{\pm\pm} = t_\pm = 0$. To make sure the singularity at $\Omega = 1/4$ is spacelike we choose $M > 1/4$. The event horizon of the black hole is at $x^+ x^- = 0$. The conformal factor is obtained from (2.6) in Kruskal coordinates as $\rho_K = 2\Omega + \ln 2 + \frac{1}{2}W_{-1}(-e^{-4\Omega})$, where W_{-1} is the product logarithm function. In Minkowski coordinates, the metric near infinity is $ds^2 = -d\sigma^+ d\sigma^-$. Normal ordering in these coordinates leads to an energy observed by asympotically inertial observers,

$$t_+(\sigma^+) = \frac{1}{4}. \tag{2.22}$$

This arises from outgoing thermal radiation flux at temperature

$$T = \frac{1}{2\pi}. \tag{2.23}$$

Using the first law $\frac{\partial S_{\mathrm{BH}}}{\partial E} = \frac{1}{T}$ and $E = \frac{NM}{12\pi}$ one finds the black hole entropy
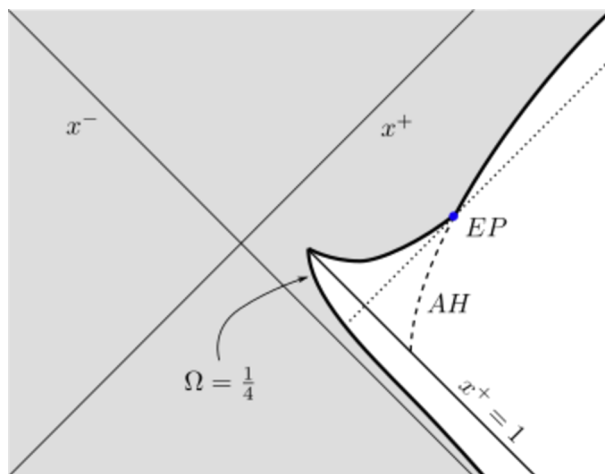
$$S_{\mathrm{BH}} = \frac{NM}{6}. \tag{2.24}$$

Figure 2. Evaporating black hole in Kruskal coordinates. The apparent horizon *AH* is on the dashed line and the evaporation endpoint is marked *EP*. The dotted line prior to *EP* is the event horizon.

2.3 Evaporating black hole in RST

A beautiful feature of the RST model is that the formation and evaporation of a black hole can be described exactly and analytically in the semiclassical limit, at least up to the endpoint of the evaporation process. The corresponding solution is

$$\Omega = -x^+x^- - \frac{1}{4}\ln(-4x^+x^-) - M(x^+ - 1)\Theta(x^+ - 1). \tag{2.25}$$

The spacetime is illustrated in figure 2 in Kruskal coordinates (2.9), and the conformal diagram appears below in figure 4. There is an incoming matter shockwave taken to be at $x^+ = 1$ (or $\sigma^+ = 0$ in Minkowski coordinates) that creates the black hole. Using the constraint equation (2.12), the matter stress tensor is

$$T_{++}^f = M\delta(x^+ - 1), \quad T_{--}^f = 0, \quad t_{\pm} = -\frac{1}{4(x^{\pm})^2}. \tag{2.26}$$

The curve $\Omega = \frac{1}{4}$ is the edge of the spacetime. This curve is initially timelike, as in the linear dilaton vacuum; here we impose reflecting boundary conditions on the matter fields. When the shockwave hits the boundary, the curve $\Omega = \frac{1}{4}$ turns spacelike, and is interpreted as the black hole singularity. After the black hole completely evaporates at the endpoint it becomes a timelike boundary again.

The solution has an apparent horizon defined by $\partial_+\Omega = 0$,[6] which is the outer boundary of the evaporating black hole. This is the curve

$$x^+(x^- + M) = -\frac{1}{4} \tag{2.27}$$

for $x^+ > 1$. The evaporation ends where the apparent horizon meets the singularity. This occurs at the endpoint denoted $P_{\rm EP}$ with

$$x_{\rm EP}^- = \frac{M}{e^{-4M} - 1}, \quad x_{\rm EP}^+ = \frac{1}{4M}(e^{4M} - 1). \tag{2.28}$$

This value of x^- also defines the event horizon.

---

[6]Interpreting Ω as the generalized quantum area, this is the boundary of the region of trapped spheres.

After the endpoint $P_{\rm EP}$, the singularity emerges from behind the horizon, the radiation continues and, problematically, the mass of the spacetime becomes arbitrarily negative. RST propose a new, modified "thunderbolt" boundary condition at $P_{\rm EP}$ which returns the system to the linear dilaton vacuum, matching along $x_{\rm EP}^-$. However, the details of this do not concern us here because all of our considerations address the behavior of the theory *prior* to the future light cone of $P_{\rm EP}$. In particular we wish to understand how much information is returned before the retarded time $x_{\rm EP}^-$.

Near $\mathcal{I}^-$, the geometry is in the linear dilaton vacuum, so the metric is $ds^2 = -d\sigma^+ d\sigma^-$ in the incoming Minkowski coordinates defined by

$$x^\pm = \pm e^{\pm\sigma^\pm}. \tag{2.29}$$

The (rescaled) initial mass of the black hole is equal to the energy of the incoming shockwave. In Minkowski coordinates

$$M = \int d\sigma^+ T_{++} = \frac{12\pi}{N} M_{ADM}, \tag{2.30}$$

where M_{ADM} is the canonical ADM mass. The initial black hole entropy is

$$S_{\rm BH} = \frac{NM}{6} = 2\pi M_{ADM}. \tag{2.31}$$

Near $\mathcal{I}^+$, the conformal factor in Kruskal coordinates is

$$e^{2\rho_K} = -\frac{1}{x^+(x^- + M)} + \mathcal{O}\left(\frac{1}{(x^+)^2}\right). \tag{2.32}$$

In the outgoing Minkowski coordinates $\tilde{\sigma}^\pm$ defined in terms of Kruskal coordinates by

$$x^+ = e^{\tilde{\sigma}^+}, \quad x^- + M = -e^{-\tilde{\sigma}^-}, \tag{2.33}$$

the metric near $\mathcal{I}^+$ takes the manifestly flat form

$$ds^2 = -d\tilde{\sigma}^+ d\tilde{\sigma}^- \left(1 + \mathcal{O}(e^{-\sigma^+})\right). \tag{2.34}$$

Comparing to (2.28) we note that the endpoint occurs at the (affine) retarded time

$$\tilde{\sigma}_{\rm EP}^- = \ln\frac{e^{4M} - 1}{M} \sim 4M - \ln M + \cdots, \tag{2.35}$$

where the corrections are suppressed by powers of e^{-4M}. To leading order the evaporation time of $4M$ matches the time it takes to radiate away a total mass M with an M-independent energy flux $t_+ = \frac{1}{4}$ in (2.22).

2.3.1 Entanglement entropy

The large N limit includes information about the quantum state of the N free scalars on $\mathcal{I}^+$. This is a pure state on any everywhere spacelike or null slice which stretches from spatial infinity to the timelike portion of the singularity — the 'origin' — where $\Omega = \frac{1}{4}$
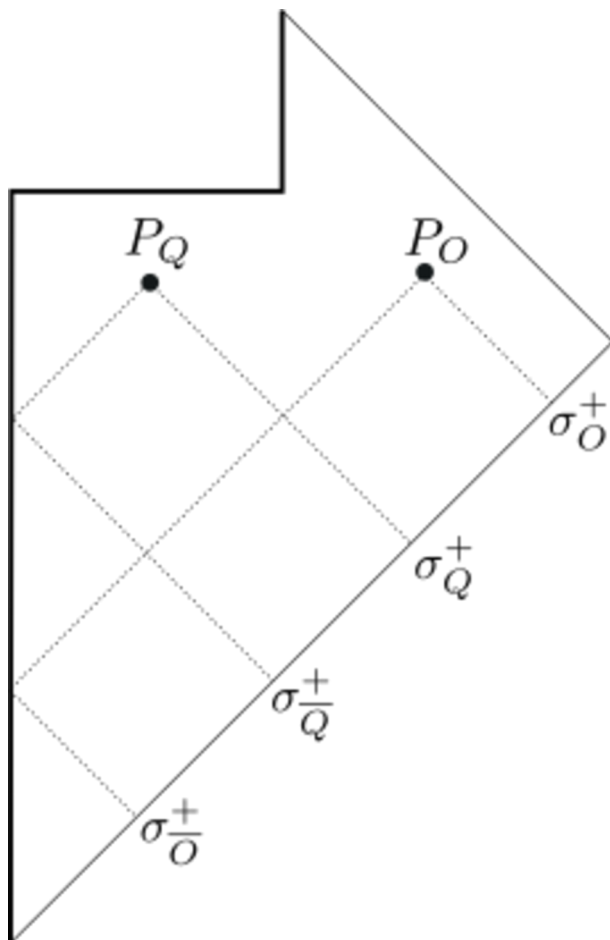
Figure 3. Points are labeled by $(\sigma_P^+, \sigma_{\bar{P}}^+)$, where $\sigma_{\bar{P}}^+$ is the value of σ^+ for the image point obtained by reflecting across the timelike boundary.

and reflecting boundary conditions are imposed on the f_k. Such slices are all prior to the future light cone of the endpoint. The quantum state on a subregion A of such a slice is a density matrix ρ_A obtained by tracing over the quantum state on the slice outside of A. The entanglement entropy

$$S_{\text{ent}}(A) = -\text{tr}\,\rho_A \ln \rho_A \tag{2.36}$$

can be computed exactly [43] when one end of A is at spatial infinity, the other at an interior point $P_O = (\sigma_O^+, \sigma_O^-)$ and the quantum fields are in their vacuum state on $\mathcal{I}^-$. P_O will often be taken to be the location of an Observer. The result is [43]

$$S_{\text{ent}}(A) = \frac{N}{6}\left[\rho_{\text{in}}(\sigma_O^+, \sigma_O^-) + \ln\frac{\sigma_O^+ - \sigma_{\bar{O}}^+}{\epsilon_{\text{uv}}}\right]. \tag{2.37}$$

Here ρ_{in} is the conformal factor evaluated in the incoming Minkowski coordinates $\sigma^\pm$ used to define the vacuum, and $\sigma_{\bar{O}}^+$ is the incoming coordinate of the null line which reflects off the boundary and arrives at P_O (see figure 3), which obeys

$$\sigma_{\bar{O}}^+ = \sigma_O^- - \ln 4. \tag{2.38}$$

$\epsilon_{\rm uv}$ is a UV regulator required by the divergent entanglement of short-wavelength modes across the point P on the spacelike slice. The factor of $\rho_{\rm in}$ appears because one wants the UV regulator to have the same fixed proper length at any slice endpoint in the spacetime.

Eq. (2.37) in fact applies to any CFT with the simple replacement of N by the central charge c. (Of course to take a large-c limit we need a sequence of theories labeled by arbitrarily large c.) A more involved formula for more general choices of the region A will be introduced below when it is needed in the definition of the island rule entropy.

If P_O is chosen to be a point on the apparent horizon, (2.37) is the entropy of the quantum fields outside the black hole. If P_O is chosen to be a point on $\mathcal{I}^+$, it is the entanglement entropy of the portion of the outgoing state prior to P_O with that after P_O. This entanglement entropy is what is physically measured by an inertial asymptotic detector that cannot detect modes shorter than $\epsilon_{\rm uv}$. In terms of the outgoing retarded Minkowski coordinate $\tilde{\sigma}_O^-$ (eq. (2.33)) of P_O, equation (2.37) becomes

$$S_{\rm ent}(\tilde{\sigma}_O^-) = \frac{N}{12}\ln(1 + Me^{\tilde{\sigma}_O^-}) + \frac{N}{6}\ln\frac{\sigma_O^+ - \sigma_{\bar{O}}^+}{\epsilon_{\rm uv}}. \tag{2.39}$$

Since $\sigma_O^+ \to \infty$ on $\mathcal{I}^+$, the second term contains both IR and UV divergences. These are independent of the retarded time of the observer and so are not associated to outgoing Hawking flux. At late times the entropy grows as $\frac{N}{12}\tilde{\sigma}_O^-$, precisely as expected for a thermal flux of N scalars at temperature $T = \frac{1}{2\pi}$. At the evaporation endpoint the net entropy in the Hawking radiation, which is the finite change in the entanglement entropy, is

$$\Delta S_{\rm ent}(\tilde{\sigma}_{\rm EP}^-) = \frac{NM}{3}. \tag{2.40}$$

This is exactly twice the initial entropy (2.31), indicating the evaporation process is not adiabatic.

For the usual reasons, it is hard to reconcile (2.40) with quantum mechanical unitarity. It appears that, to the extent that it is well-defined, the RST model (and its cousins) destroys information.[7] Hence we seek some kind of modified semiclassical theory with a different entropy flux. This brings us to the next section.

3   The island rule for an evaporating black hole

3.1   Quantum extremal surface

The island rule associates an entropy to a point P_O which involves an extremization of a function over the location P_Q of the quantum extremal surface. The formula for the island rule entropy is [2–4]

$$S_I(P_O) = \min {\rm ext}_{P_Q} \left[S_{\rm gen}(I \cup R) \right] . \tag{3.1}$$

---

[7]In general there is a concern (or hope) that off-diagonal terms in the Hawking radiation density matrix could conspire to correct the leading semiclassical (or large-N) result. This is clearly impossible in the RST model — which is by definition on a fixed topology — due to the flavor symmetry of the N bosons. They are strictly decoupled and therefore the flavor information cannot be encoded in the off-diagonal elements of the density matrix.
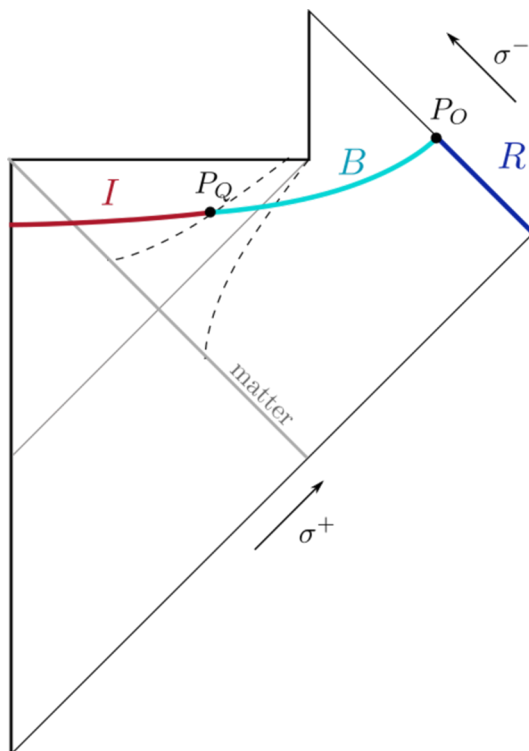
Figure 4. Island in the evaporating black hole. The Observer at P_O collects the radiation in region R. The island is I, and $I \cup B \cup R$ is a Cauchy slice. The endpoint of the island is the quantum extremal surface P_Q. The dashed lines are the apparent horizon and the curve of extremal surfaces.

P_Q can consist of several points in general, but in this section I is an "island" region which extends from the timelike origin to a single point P_Q, and R is the region from spacelike infinity to P_O through which Hawking radiation passes (see figure 4). We also use B to denote the region between P_Q and P_O. We restrict the points so that $I \cup B \cup R$ is a complete everywhere spacelike or null Cauchy surface. The generalized entropy is

$$S_{\text{gen}}(I \cup R) = S_{\text{grav}}(P_Q) + S_{\text{ent}}(I \cup R) . \tag{3.2}$$

The first term S_{grav} is the Bekenstein-Hawking entropy. In the RST model this is [43]

$$S_{\text{grav}} = \frac{N}{6}\left(e^{-2\rho_K} - \frac{\rho_K}{2}\right) + \frac{N}{24}(\ln 4 - 1) + \frac{N}{6}\ln\epsilon_{\text{uv}} , \tag{3.3}$$

where ρ_K is in Kruskal coordinates.[8] The $\ln\epsilon_{\text{uv}}$ term in (3.3) will cancel against the UV divergence of the matter entropy coming from the point P_Q when computing S_{gen}. The point P_Q which extremizes the right hand side of (3.1) is the quantum extremal surface. If there are multiple extrema the generalized entropy is computed by picking the solution which minimizes the generalized entropy. The S_I defined in (3.1) is a natural proposed generalization to the asymptotically flat 2D context of formulae which can be derived in

---

[8]We note $\rho_{\text{in}} = \rho_K + \frac{1}{2}(\sigma^+ - \sigma^-)$.

some AdS contexts, but it has not been identified as an asymptotic observable quantity in the standard formulation of the RST model. One may seek to modify the RST model so that (3.1) is a physical observable. This would be a major achievement.

The matter term $S_{\text{ent}}(I \cup R) = S_{\text{ent}}(B)$ is given by a formula which is more complicated than (2.37) because B does not terminate at spatial infinity. Assuming the scalars are compactified at the self-dual radius, one finds [49]

$$S_{\text{ent}}(I \cup R) = \frac{N}{6} \ln \left[\frac{d^2(P_O, P_Q) d(P_O, P_{\bar{O}}) d(P_Q, P_{\bar{Q}})}{\epsilon_{\text{uv}}^2 d^2(P_Q, P_{\bar{O}}) e^{-\rho_{\text{in}}(P_Q) - \rho_{\text{in}}(P_O)}} \right] \tag{3.4}$$

with $d^2(P, P') \equiv (\sigma_P^+ - \sigma_{P'}^+)(\sigma_{P'}^- - \sigma_P^-)$ the Minkowski distance. When the conformal cross ratio of the four endpoints satisfies $\frac{d(P_Q, P_O) d(P_{\bar{O}}, P_{\bar{Q}})}{d(P_Q, P_{\bar{O}}) d(P_{\bar{Q}}, P_O)} \ll 1$, this can be approximated by

$$S_{\text{ent}}(I \cup R) = \frac{N}{6} \ln \frac{-(\sigma_{\bar{Q}}^+ - \sigma_{\bar{O}}^+)(\sigma_Q^+ - \sigma_O^+)}{\epsilon_{\text{uv}}^2 e^{-\rho_{\text{in}}(P_Q) - \rho_{\text{in}}(P_O)}} \ . \tag{3.5}$$

In fact, this latter equation applies to any compact CFT by an OPE argument [50]. We are assuming the scalar is compact (with $O(1)$ radius) to avoid subtleties with the zero mode that would modify this formula [43, 49].

This is the CFT entropy in the vacuum state. To describe a shockwave we may use a coherent state of the matter fields which does not affect the entropy [43], so this formula also applies to the evaporating black hole.

3.2 Evaporating black hole

In this subsection we compute the island rule entropy for the case of an evaporating black hole. We will focus on the regime where the entanglement entropy is given by (3.5), since it can be checked that the this is always valid for the island. Adding (3.3) to (3.5) and sending $P_O \to \mathcal{I}^+$ one finds

$$S_{\text{gen}}(I \cup R) = \frac{N}{6} \left[\Omega(P_Q) - \frac{1}{4} + \frac{1}{2}(\sigma_Q^+ - \sigma_{\bar{Q}}^+) + \ln(\sigma_{\bar{Q}}^+ - \sigma_{\bar{O}}^+) \right] \tag{3.6}$$

$$+ \frac{N}{6} \ln \left(\frac{\sigma_O^+}{\epsilon_{\text{uv}} \sqrt{1 - 4M e^{\sigma_{\bar{O}}^+}}} \right) .$$

Note that the UV and IR divergences $\ln \frac{\sigma_O^+}{\epsilon_{\text{uv}}}$ from the point P_O remain. However they do not affect the extremization. In fact only the first line depends on the point P_Q. The extremality conditions $\partial_{\sigma_{\bar{Q}}^+} S_{\text{gen}} = \partial_{\sigma_Q^+} S_{\text{gen}} = 0$ are

$$M(\sigma_{\bar{Q}}^+ - \sigma_{\bar{O}}^+) = e^{-\sigma_Q^+} = 4M - e^{-\sigma_{\bar{Q}}^+} \ . \tag{3.7}$$

In Kruskal coordinates,

$$x_Q^+ (x_Q^- + M) = \frac{1}{4} \, , \tag{3.8}$$

$$x_O^- = x_Q^- e^{\frac{4(x_Q^- + M)}{M}} . \tag{3.9}$$

A solution to these equations defines a quantum extremal surface (QES). In this case there are two solutions depending on the chosen branch cut:

$$x_Q^- = \frac{M}{4} W_{-1}\left(\frac{4x_O^-}{e^4 M}\right) , \qquad x_Q^+ = \frac{1}{4(M + x_Q^-)} ; \tag{3.10}$$

$$x_Q^- = \frac{M}{4} W_0\left(\frac{4x_O^-}{e^4 M}\right) , \qquad x_Q^+ = \frac{1}{4(M + x_Q^-)} . \tag{3.11}$$

$W_k(x)$ is the product logarithm function with branch cut indicated by k.

Once we have found the QES, and solved for $\sigma_Q^\pm$, $S_I(P_O)$ in (3.1) becomes a function only of the retarded observer time $\tilde{\sigma}_O^-$. The explicit expression however involves the product logarithm function and is unilluminating. A simple formula for large mass is given below.

Recall that the apparent horizon is on the curve $x^+(x^- + M) = -\frac{1}{4}$, and the event horizon is slightly before $x^- + M = 0$. Therefore the QES sits inside the event horizon, and lies on a curve obtained by reflecting the apparent horizon across the event horizon.

Because all time foliations are on equal footing in general relativity, one generically expects that there is no canonical way to associate a boundary or retarded time with a point in the interior — especially if the interior is inside the black hole. Contrary to this expectation, (3.9) explicitly associates a retarded time to the QES inside the black hole. Such an association might seem required in any mechanism which pulls information out of the black hole.

3.2.1   When does the QES hit the singularity?

Eventually the QES always collides with the singularity. After this point, we do not know how to define the island rule, and the calculation breaks down. The collision occurs where the QES curve given by the first equation in (3.8) hits the singularity $\Omega = \frac{1}{4}$, namely

$$x_{QE}^- = -\frac{M}{1 + e^{2-4M}} , \tag{3.12}$$

which is greater than x_{EP}^-. Using the map (3.9) between QES time and retarded time and taking M to be large, the collision 'appears' to an observer on $\mathcal{I}^+$ at retarded time

$$\tilde{\sigma}_{OQE}^- \approx 4M - \ln M - \ln 3 - 2 + \cdots \tag{3.13}$$

where the corrections are suppressed by powers of e^{-4M}. Comparing to expression (2.35) we see that the collison appears (for large mass) before the endpoint by an amount

$$\Delta\tilde{\sigma}^- = -2 - \ln 3. \tag{3.14}$$

The conclusion is that even for large mass, the breakdown occurs only a time of order one before the endpoint. We do not know how to extend the analysis beyond this time, but this will not prevent us from reproducing the Page curve to leading order at large M.

One may check directly at order M that the island rule entropy vanishes when the QES hits the singularity. Interestingly, there is no order M gravitational contribution (S_{grav}) to

the island rule entropy, because the end of the island is on the singularity where the area is small. Therefore the entire expression can be interpreted as CFT entanglement entropy of $I \cup R$ or equivalently B. The entanglement of the projection of region B back to $\mathcal{I}^-$ is represented by the $\ln(\sigma_Q^+ - \sigma_O^+)$ term in (3.6). This projected region is a narrow sliver of affine length e^{-4M}, due to the tremendous blueshift. Therefore this term gives a *negative* contribution of $-\frac{2NM}{3}$ to the entanglement entropy. This is possible because we have subtracted the cutoff dependence, which is operationally equivalent to taking $\epsilon_{\mathrm{uv}} = 1 = \lambda$. The (subtracted) entanglement of a region smaller than the cutoff then becomes negative. However this negative contribution is exactly cancelled by the $\frac{N\rho_{\mathrm{in}}(\sigma)}{6}$ factors at the two ends of the interval B (the third and last factors in (3.6)), each of which give a factor of $\frac{NM}{3}$.

Of course the combined region $I \cup B \cup R$ also has zero entanglement entropy since it is a Cauchy slice. It is curious that removing the region B does not change the entanglement entropy. This is a very special property of the exact point σ_{OQE}^- where the QES is seen to hit the singularity. Had we waited a time of order one the $\frac{N\rho_{\mathrm{in}}(\sigma)}{6}$ factors would not have changed much, but the projected interval on $\mathcal{I}^-$ would have become exponentially smaller, resulting in a negative entanglement entropy of order M. This provides a stringent consistency test of the island rule.

3.3 Page curve

In this subsection we derive the Page curve to leading order at large M, dropping subleading terms of order $\ln M$.

According to the island rule (3.1), the entropy as a function of $\tilde{\sigma}_O^-$ is obtained by first finding all surfaces obeying the quantum extremality condition, then choosing the one with minimal entropy. One choice is the trivial surface, with no island. The island rule entropy in this case simply equals the quantum entanglement entropy $S_{\mathrm{ent}}(\tilde{\sigma}_O^-)$ in equation (2.39), which follows the Hawking curve. For a single island, the only other choice is (3.10). Therefore the island rule entropy is

$$S_I(\tilde{\sigma}_O^-) = \min\left[S_{\mathrm{ent}}(\tilde{\sigma}_O^-),\ S_{\mathrm{gen}}(I \cup R)\right] \tag{3.15}$$

where the second term is evaluated at the extremum given by (3.10).

For M large and $\tilde{\sigma}_O^-$ of order M we can solve to leading order for the location of the QES using (3.8) and (3.9). One finds

$$x_Q^+ \approx \frac{3}{4}e^{\tilde{\sigma}_O^-}, \qquad x_Q^- + M \approx \frac{1}{3}e^{-\tilde{\sigma}_O^-}\ . \tag{3.16}$$

Plugging into the island rule one finds

$$S_I = \min \frac{N}{24}\left[2\tilde{\sigma}_O^-,\ \tilde{\sigma}_{\mathrm{EP}}^- - \tilde{\sigma}_O^-\right] + \frac{N}{6}\ln\frac{\sigma_O^+}{\epsilon_{\mathrm{uv}}} \tag{3.17}$$

where $\tilde{\sigma}_{\mathrm{EP}}^- \approx 4M$ is the evaporation endpoint.

The change in the entropy (which subtracts the IR and UV divergences), plotted in figure 5, has the form anticipated by Page [51, 52]: it grows linearly until an O(1) fraction of
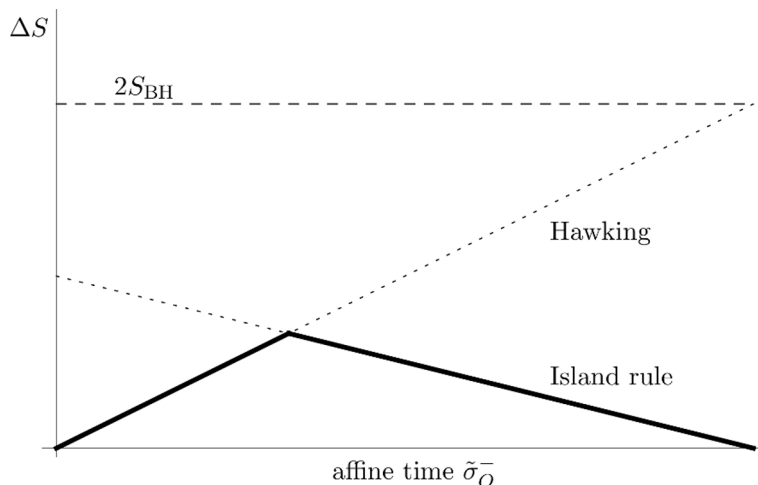
Figure 5. Page curve (assuming a large initial mass $M \gg 1$).

the black hole has evaporated, then decreases back to zero. The Page time $t_P = \frac{4M}{3}$, when the entropy starts to decrease, is at the point where 1/3rd of the black hole has evaporated. This is because black hole evaporation is non-adiabatic, and in the RST model, entropy decrease $-\delta S$ of the black hole produces an entropy $+2\delta S$ in the radiation. Therefore the black hole entropy is equal to the radiation entropy after 1/3rd of the black hole evaporates.

3.4 Scrambling time

We now compute the scrambling time associated to the island rule entropy following [2, 3]. The scrambling time is the time it takes for information dropped into a black hole to reappear in the Hawking radiation. The island rule effectively asserts that the information in the island at time x_Q^- is available at $\mathcal{I}^+$ at the canonically associated retarded time x_O^- in (3.9). The dropped information hence becomes available when it hits the QES curve. The main difference in asymptotically flat space compared to AdS is that we need to keep track of the initial time in the scrambling experiment. The result agrees with Hayden and Preskill [53] (see also [54, 55]), namely $t_{\rm scr} \sim \frac{\beta}{2\pi}\ln\frac{S}{N} \sim \ln(M)$ for large N and M.

The experiment is illustrated in figure 6. After the Page time, an object is dropped into the black hole by a distant observer, who is at rest in the inertial coordinates near infinity,

$$\tilde{\sigma}^+ = t + y, \quad \tilde{\sigma}^- = t - y . \tag{3.18}$$

The observer should be far from the black hole, but close enough so the black hole is effectively stationary for the whole experiment. The observer collects the Hawking radiation along $\mathcal{I}^+$. Let $\tilde{\sigma}_0^-$ be the retarded time when the object reaches a few Schwarzschild radii from the black hole (the precise definition is given below) and $\tilde{\sigma}_{OQ}^-$ be the retarded time associated via (3.9) to the object crossing the QES curve and entering the island. Then
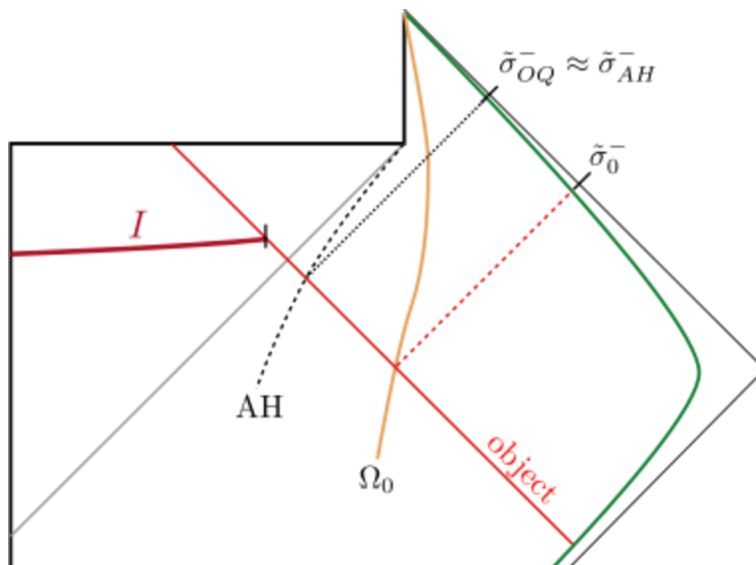
Figure 6. Scrambling experiment.

the scrambling time is defined by[9]

$$t_{\text{scr}} = \tilde{\sigma}_{OQ}^- - \tilde{\sigma}_0^- \ . \tag{3.19}$$

To see that this definition makes sense, note that the experiment has three stages: first the object falls to near the black hole in time t_{fall}, then it is scrambled, then the Hawking radiation carrying information about the object takes the same time t_{fall} to return from the near region to the observer. Thus the total proper time elapsed at infinity, from dropping the object to seeing it return in the Hawking radiation, should be

$$\Delta t = t_{\text{scr}} + 2t_{\text{fall}} \ . \tag{3.20}$$

Since the black hole is effectively stationary, we can see that this definition of the scrambling time implies (3.19) upon using $\tilde{\sigma}_0^- = t_{\text{obj}} + 2t_{\text{fall}} - y$.

Now we need to define 'near' the black hole. In higher dimensions this would be a region within a few Schwarzschild radii. In two dimensions, the coordinate-invariant analogue is to define the near region as

$$\Omega \lesssim \Omega_0 \equiv (1 + A)\Omega_H \, , \tag{3.21}$$

where Ω_H is Ω at the horizon, and $A > 0$ is an $O(1)$ constant. We want A to be large so that we can neglect black-hole-related redshifts, but to be parametrically small compared to N or M. Thus we define $\tilde{\sigma}_0^-$ to be the retarded time when the object reaches the curve $\Omega = \Omega_0$. This is when the observer along $\mathcal{I}^+$ starts the timer to define the scrambling time.

---

[9]Note that light rays emitted by the object when it crosses the apparent horizon hover near the black hole for a long time and suffer a great delay before arriving at infinity. Hence we cannot equate the scrambling time with the difference between this retarded crossing time $\tilde{\sigma}_{AH}^-$ and $\tilde{\sigma}_{OQ}^-$. In fact that difference equals $-\ln 3$ and is negative! See [2–4, 56] for discussion.

We take the object to fall on a null trajectory with $x^+ = x^+_{\text{obj}}$. We can treat the black hole as stationary, with

$$\Omega \approx -x^+(x^- + M) + \frac{6S_{\text{BH}}}{N} \; . \tag{3.22}$$

where as we have seen $\frac{6S_{\text{BH}}}{N} = M$. The horizon is at $x^- = -M$, i.e. $\Omega = \frac{6S_{\text{BH}}}{N}$, so we have $\Omega_0 = (1 + A)\frac{6S_{\text{BH}}}{N}$. The object reaches the curve $\Omega = \Omega_0$ at

$$\tilde{\sigma}^-_0 = \ln\left(\frac{Nx^+_{\text{obj}}}{6AS_{\text{BH}}}\right) \; . \tag{3.23}$$

According to (3.16), the object is seen at $\mathcal{I}^+$ to enter the island at the retarded time

$$\tilde{\sigma}^-_{OQ} = \ln\frac{4x^+_{\text{obj}}}{3}. \tag{3.24}$$

Therefore

$$\tilde{\sigma}^-_{OQ} = \tilde{\sigma}^-_0 + \ln\left(\frac{8AS_{\text{BH}}}{N}\right) \; . \tag{3.25}$$

The $\ln(8A)$ term is subleading, so for the scrambling time we find to leading order

$$t_{\text{scr}} = \frac{\beta}{2\pi}\ln\frac{S_{\text{BH}}}{N} \; , \tag{3.26}$$

where the inverse temperature of the black hole is $\beta = 2\pi$. This agrees with [53, 54], including the $\ln N$ correction also found in [2].

4 The eternal black hole

In this section we apply the island rule to the eternal black hole in the RST model. In addition to being mathematically simpler than the evaporating black hole, the late time behavior of the quantum extremal surfaces in the evaporating black hole is well-approximated by those in the eternal black hole. Moreover, as discussed in [43, 56, 57], the eternal black hole furnishes a simple version of the information paradox: at late times, the black hole in the Hartle-Hawking vacuum seems to have an arbitrarily large entanglement with the Hawking radiation, despite the fact that it has a fixed, finite Bekenstein-Hawking entropy. Our discussion here follows a closely related AdS version in [56].

Consider the entropy in the region $R \cup \overline{R}$, where R is near $\mathcal{I}^+$ and $\overline{R}$ is its reflection on the other side of the Penrose diagram. (See figure 7.) Both intervals are of equal size, and as we make them bigger the entanglement entropy increases without bound. This can be seen as follows. Since the global quantum state is pure we consider the entropy on the complement, which is a single interval extending from the point $\overline{P}_O$ on the left to P_O on the right, with both points near $\mathcal{I}^+$. The entanglement entropy of a conformal field theory on an interval $[\bar{x}_O, x_O]$ in the spacetime $ds^2 = -e^{2\rho_K}dx^+dx^-$ is

$$S_{\text{ent}}(R \cup \overline{R}) = \frac{N}{6}\ln\left[\frac{(x^+_O - \bar{x}^+_O)(\bar{x}^-_O - x^-_O)}{\epsilon^2_{\text{uv}}e^{-\rho_K(x_O) - \rho_K(\bar{x}_O)}}\right]. \tag{4.1}$$
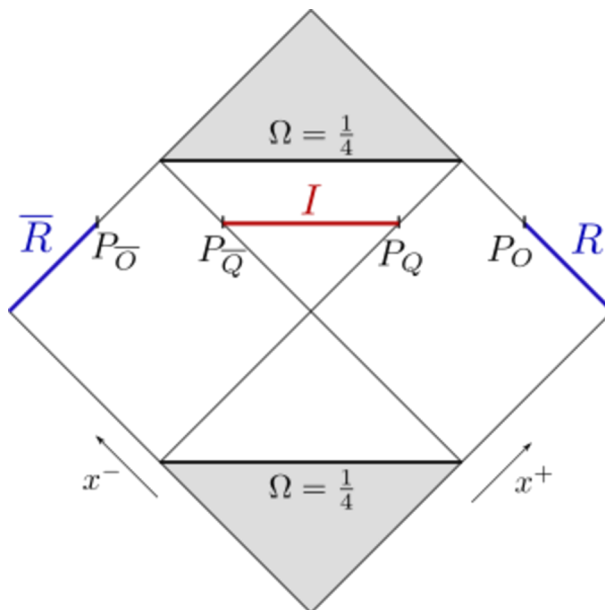
Figure 7. Island in the eternal black hole. With the radiation collected on $\mathcal{I}^+$, the endpoints of the island are on the horizon.

In the rest of this section, we set $\sigma_O^\pm = t_O \pm y_O$, expand in the regime $M \gg 1, t_O \gg 1, y_O \gg \ln M, \frac{1}{\epsilon_{\mathrm{uv}}} \gg 1$, and drop terms of $O(1)$. Considering the interval of interest in the black hole background, the entanglement entropy becomes

$$S_{\mathrm{ent}}(R \cup \overline{R}) \approx \frac{N}{3} t_O - \frac{N}{3} \ln \epsilon_{\mathrm{uv}} \,. \tag{4.2}$$

This increases linearly with time. On the other hand, unitarity suggests[10] that this be less than the entropy of the black hole together with the thermal bath on $B \cup \overline{B}$, where B is the region between the black hole and R:

$$S(R \cup \overline{R}) \lesssim 2\left(S_{\mathrm{grav}} + S_{\mathrm{ent}}(B)\right) \,, \tag{4.3}$$

with the gravitational term evaluated at the horizon. This is the entropy of the black hole plus bath, with a factor of 2 to account for the two sides. The right-hand side of (4.3) is time-independent, so we can calculate it at $t = 0$, taking B to be the region from the bifurcation point to y_O. Then

$$S_{\mathrm{grav}} + S_{\mathrm{ent}}(B) \approx \Omega_H + \frac{N}{6} \ln\left[\frac{-x_O^+ x_O^-}{\epsilon_{\mathrm{uv}} e^{-\rho_K(P_O)}}\right] \tag{4.4}$$

$$\approx \frac{NM}{6} + \frac{N y_O}{6} - \frac{N}{6} \ln \epsilon_{\mathrm{uv}} \,, \tag{4.5}$$

---

[10]The entanglement entropy (4.2) could only be measured by metaobservers who access both asymptotic regions of the eternal geometry. Hence the paradox here is less physical than, though a close mathematical analogy to, the one encountered in the previous section or discussed in [43, 57].

up to subleading terms not proportional to M or y_O. Thus the expected unitarity bound (4.3) is

$$S(R \cup \overline{R}) \lesssim \frac{NM}{3} + \frac{Ny_O}{3} - \frac{N}{3} \ln \epsilon_{\text{uv}} .$$ (4.6)

There may be subleading corrections to this bound, but in any case eventually it will be overtaken by the linear growth in (4.2). Hence the explicit expression (4.2) clearly violates (4.3).

Now we consider the island rule entropy. We are looking for a quantum extremal surface which is two points bounding an island containing the interior of the black hole. Then the matter entanglement is computed on the surface $\overline{R} \cup I \cup R$ where I is the island. By purity of the global state we can again consider the entropy on the complement region, which is two intervals $B \cup \overline{B}$. At late times these two intervals are far apart and the total entanglement entropy becomes twice the single interval answer. We then have

$$S_{\text{ent}}(\overline{R} \cup I \cup R) = \frac{N}{3} \ln \left[\frac{(x_O^+ - x_Q^+)(x_Q^- - x_O^-)}{\epsilon_{\text{uv}}^2 e^{-\rho_K(P_O) - \rho_K(P_Q)}} \right] .$$ (4.7)

The gravitational entropy also has two contributions, so is given by twice the formula (3.3) evaluated at the point x_Q. The generalized entropy is

$$S_{\text{gen}}(\overline{R} \cup I \cup R) = S_{\text{grav}} + S_{\text{ent}}$$ (4.8)
$$= \frac{N}{3} \left(\Omega(P_Q) + \ln \left[\frac{(x_O^+ - x_Q^+)(x_Q^- - x_O^-)}{\epsilon_{\text{uv}}} \right] + \rho_K(P_O) + \ln 2 - \frac{1}{4} \right) .$$

Extremizing with respect to $x_Q^{\pm}$ gives $x_Q^{\pm} \approx -\frac{1}{x_O^{\mp}}$ at large x_O^+. In this approximation we find that the endpoint of the island lies on the horizon at

$$\sigma_Q^+ = \sigma_O^- .$$ (4.9)

The entropy becomes

$$S_{\text{gen}} \approx 2 S_{\text{BH}} + \frac{Ny_O}{3} - \frac{N}{3} \ln \epsilon_{\text{uv}} .$$ (4.10)

Comparing to the case with no island (4.2) shows that the island saddle dominates for

$$t_O \gtrsim M + y_O$$ (4.11)

After the Page transition, the entropy saturates the unitarity bound (4.6) up to subextensive terms.

Acknowledgments

A   Local analysis of replica wormholes

For black holes in AdS$_2$, the island rule has been derived directly from the gravitational path integral [5, 6]. The path integral is used to calculate the replica partition functions

$$Z_n = \text{Tr} \left(\rho_A \right)^n \ . \tag{A.1}$$

This can be analytically continued to non-integer n by continuing the solution itself, and used to obtain the von Neumann entropy,

$$S(\rho_A) = -\partial_n \left(\frac{Z_n}{(Z_1)^n} \right)_{n=1} \ . \tag{A.2}$$

It is not obvious whether these results carry over to asymptotically flat spacetime, where gravity is dynamical everywhere. In this appendix we make a step in this direction. We discuss replica wormhole geometries for the eternal black hole in the RST model, construct the solutions locally near the quantum extremal surface, and show that the equations of motion enforce the QES condition. However we will not construct the wormhole saddles globally, so this is not a complete derivation of the island rule.

The local analysis closely follows the derivation of the Ryu-Takayanagi formula by Lewkoywycz and Maldacena [58] and the general analysis of quantum extremal surfaces by Dong and Lewkowycz [59] (see also [60]). By assuming a large-N matter sector, we sidestep some of the difficulties in defining the matter entanglement entropy in [59].

A.1   Replica geometries

We will focus on the eternal black hole at $t = 0$. The same results apply (locally near the QES) to any spacetime with a time-reflection symmetry, such that it can be prepared by a Euclidean path integral.

Take R to be the semi-infinite region $y \in [y_O, \infty)$, with $y_O \gg 1$, on the right side of the Penrose diagram for the eternal black hole. This is analogous to the region where an observer collects the Hawking radiation, but it is at $t = 0$. Its reflection across the Penrose diagram is $\bar{R}$. We are interested in the von Neumann entropy of $R \cup \bar{R}$, and the associated replica partition functions

$$Z_n = \text{Tr} \left(\rho_{R \cup \bar{R}} \right)^n , \tag{A.3}$$

calculated from the gravitational path integral.

Consider first $n = 1$. The partition function $Z_1 = \text{Tr} \, \rho_{R \cup \bar{R}}$ is independent of the region. The path integral has asymptotic boundary conditions specifying that we have an S^1 of size 2π, and the dilaton asymptotics given by (2.21).[11] The saddlepoint is simply the Euclidean black hole, i.e., the cigar geometry:


---
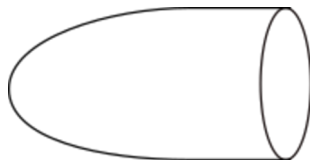
[11]The canonical ensemble is ill-defined in the RST model, since the temperature is not a free parameter. We therefore fix both the leading and subleading terms in Ω. As usual, boundary terms and counterterms are needed to make the action finite, but we will not need the details.

Next we turn to $n = 2$. The standard replica manifold has the same topology as two copies of the Euclidean black hole, glued along the regions R and $\bar{R}$:
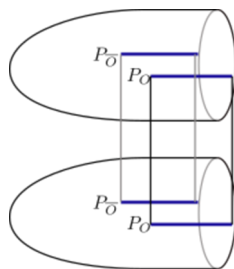

Each cigar is associated to a factor of $\rho_{R \cup \bar{R}}$, and the gluing accounts for the matrix multiplication in $\mathrm{Tr}\,(\rho_{R \cup \bar{R}})^2$. The gluing is defined cyclically, so the top of region R in copy 1 is glued to the bottom of region R in copy 2, and vice-versa. Note that $\bar{R}$ is half way around the thermal cycle from R. There are two distinct asymptotic S^1 boundaries. To find the saddle, the equations of motion should be solved everywhere except at the branch points P_O and $P_{\bar{O}}$, where there are 2π conical excesses.

Still at $n = 2$, we can also consider higher topology, contributions to the path integral, connected in the interior of the geometry:
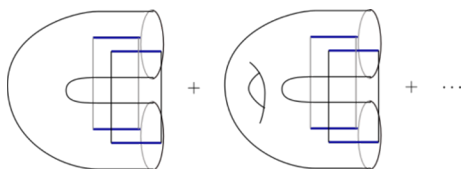
We impose the same boundary conditions as before — two copies of S^1 and the appropriate dilaton asymptotics, with 2π conical excesses at P_O and $P_{\bar{O}}$.[12] The interior of the manifold, including the wormhole connecting the two replicas, is completely smooth by the equations of motion.

Proceeding to higher n, we can in principle consider all topologies at each n. The corresponding saddles, if they exist, are the replica wormholes. We will restrict to manifolds with a $\mathbf{Z}_n$ cyclic symmetry permuting the n copies of the cigar. The geometry of a $\mathbf{Z}_n$-symmetric replica wormhole can be represented as a branched cover of a single cigar, with an even number of n-fold twist defects:

---
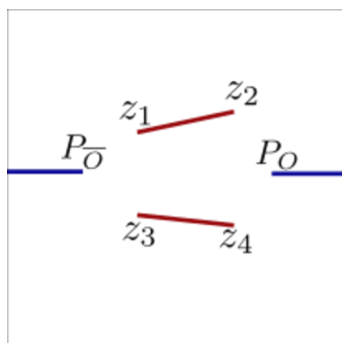
[12]Away from $n = 1$ there is an ambiguity in where we place the point P_O, because it is in a different metric. This ambiguity would need to be fixed to fully specify the replica saddles, but it will not affect our discussion near $n \sim 1$ as it only appears at $O((n-1)^2)$.

Here we have switched to a planar coordinate, so the metric is asymptotically $dzd\bar{z}/(z\bar{z})$, the thermal cycle is $z \to ze^{2\pi i}$, and the horizon (for $n = 1$) is at $z = 0$. The full manifold $\mathcal{M}_n$ has n copies of this picture, glued cyclically along all of the cuts. This full manifold is smooth everywhere except at the *original* branch points P_O and $P_{\bar{O}}$, where it has $2\pi(n-1)$ conical excesses. In particular it is smooth at the new, dynamical branch points z_i. Here the equations of motion impose

$$ds^2 \approx \frac{1}{n^2}|z - z_i|^{2/n-2}dzd\bar{z} \quad \text{as} \quad z \to z_i \ . \tag{A.4}$$

That is, $ds^2 \approx dwd\bar{w}$ in the coordinate $w = (z - z_i)^{1/n}$ that covers a neighborhood of the defect, so the full manifold is smooth at this point, as required.

The Euclidean constraint equations of the RST model in conformal gauge $ds^2 = e^{2\rho}dzd\bar{z}$ are

$$\left(e^{-2\phi} + \frac{N}{48}\right)(4\partial\rho\partial\phi - 2\partial^2\phi) - \pi T_{zz} = 0 \, , \tag{A.5}$$

and similarly with bars. Here T_{zz} is the full stress tensor, including the anomaly. The metric dependence can be extracted by writing

$$T_{zz} = T_{zz}^{\text{flat}} - \frac{N}{12\pi}(\partial\rho\partial\rho - \partial^2\rho) \, , \tag{A.6}$$

where T_{zz}^{flat} is the stress tensor in the metric $dzd\bar{z}$. T_{zz}^{flat} is the matter stress tensor on the replica manifold with metric $dzd\bar{z}$. It is therefore the same stress tensor that would appear in a purely CFT calculation of the Renyi entropy, without gravity. The other equations of motion are

$$\partial\bar{\partial}\Omega = \partial\bar{\partial}\chi = e^{2(\chi-\Omega)} \ . \tag{A.7}$$

ρ is regular near the original defects P_O and $P_{\bar{O}}$. According to (A.4), near the dynamical defects, it obeys

$$\rho \approx \left(\frac{1}{n} - 1\right)\ln|z - z_i| \quad \text{as} \quad z \to z_i \ . \tag{A.8}$$

To summarize, a $\mathbf{Z}_n$-symmetric replica wormhole is a solution to the equations (A.5) and (A.7), subject to (A.8) at the defects, with the Euclidean black hole boundary conditions at infinity. The fields are also required to be single-valued on the quotient manifold, i.e., on the z-plane. As we will see, both the positions of the defects z_i and the behavior of ϕ near these points are fixed by the EOM.

Note that this formulation makes sense for non-integer n. This is the advantage of working on the quotient space.

A.2 Derivation of the QES

We will not attempt to solve the full replica equations in this paper, but we will solve them locally near the defects, for $n \sim 1$. We will see that this fixes the z_i to sit at extrema of the generalized entropy.

Let us shift the coordinates so that a dynamical defect sits at $z = 0$. We will solve the constraint equation (A.5) (and its barred version) near $z \sim 0$ to leading order in $n - 1$. Expand

$$\rho \to \rho + (n-1)\delta\rho, \quad \phi \to \phi + (n-1)\delta\phi, \quad T^{\text{flat}}_{\mu\nu} \to T^{\text{flat}}_{\mu\nu} + (n-1)\delta T^{\text{flat}}_{\mu\nu} . \tag{A.9}$$

The boundary condition (A.8) near the defect is

$$\delta\rho \sim -\frac{1}{2}\ln(z\bar{z}) . \tag{A.10}$$

The stress tensor $T^{\text{flat}}_{\mu\nu}$ is the same stress tensor that appears in the CFT calculation of entanglement entropy in flat space, without gravity [61, 62]. It satisfies a conformal Ward identity that relates its singular behavior to the CFT entanglement entropy [62],

$$2\pi\delta T^{\text{flat}}_{zz} = \frac{N/12}{z^2} - \frac{\partial S^{\text{flat}}_{\text{CFT}}}{z} + \text{reg.} \tag{A.11}$$

$$2\pi\delta T^{\text{flat}}_{\bar{z}\bar{z}} = \frac{N/12}{\bar{z}^2} - \frac{\bar{\partial} S^{\text{flat}}_{\text{CFT}}}{\bar{z}} + \text{reg.} \tag{A.12}$$

In the $1/z^2$ term we used the chiral scaling dimension of the twist operator,

$$h_n = \bar{h}_n = \frac{N}{24}(n - 1/n) . \tag{A.13}$$

$S^{\text{flat}}_{\text{CFT}}$ is the matter entanglement entropy in the metric $dz d\bar{z}$. Note that it depends implicitly on the global structure of the replica manifold. In particular it depends on the locations of P_O, $P_{\bar{O}}$, and all of the dynamical defects z_i.

Now we will show that the constraints near $z \to 0$ fix the position of the defect. The dilaton has an expansion

$$\delta\phi = a_{00} + a_{10}z + a_{01}\bar{z} + a_{11}z\bar{z} + a_{20}z^2 + a_{02}\bar{z}^2 + \cdots \tag{A.14}$$
$$\ln(z\bar{z})(b_{00} + b_{10}z + b_{01}\bar{z} + b_{11}z\bar{z} + b_{20}z^2 + b_{02}\bar{z}^2 + \cdots)$$

Note that terms like $z\ln z$ are not allowed, except in the combination $z\ln(z\bar{z})$. This would not be single valued on the replica manifold. Plug this expansion into (A.5) and expand $z, \bar{z} \to 0$. The leading quadratic singularity fixes b_{00}. The simple pole term in the zz constraint fixes b_{01} and b_{10}. The we plug the results into the $\bar{z}\bar{z}$ constraint and from the simple pole find the conditions

$$\partial(S_{\text{BH}} + S_{\text{CFT}}) = \bar{\partial}(S_{\text{BH}} + S_{\text{CFT}}) = 0 . \tag{A.15}$$

Here

$$S_{\text{BH}} = 2e^{-2\phi} - \frac{N}{12}\phi + \text{const.} \tag{A.16}$$

is the gravitational contribution to the entropy [43, 63], and

$$S_{\text{CFT}} = S^{\text{flat}}_{\text{CFT}} + \frac{N}{6}\rho(P_{QES}) + \dots \tag{A.17}$$

This last equation is the Weyl transformation of the entropy from $dzd\bar{z} \to e^{2\rho}dzd\bar{z}$. The dots refer to the Weyl factors at all other twist points, which do not affect the extremization (A.15).

Thus we have derived the position of the QES directly from the replica equations of motion. It would be very interesting to find global solutions in Lorentzian signature, with P_O and $P_{\bar{O}}$ taken to $\mathcal{I}^+$.

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

References

[1] G.W. Gibbons and S.W. Hawking, *Action integrals and partition functions in quantum gravity*, *Phys. Rev. D* **15** (1977) 2752 [INSPIRE].

[2] G. Penington, *Entanglement wedge reconstruction and the information paradox*, arXiv:1905.08255 [INSPIRE].

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

[4] A. Almheiri, R. Mahajan, J. Maldacena and Y. Zhao, *The Page curve of Hawking radiation from semiclassical geometry*, *JHEP* **03** (2020) 149 [arXiv:1908.10996] [INSPIRE].

[5] A. Almheiri et al., *Replica wormholes and the entropy of Hawking radiation*, *JHEP* **05** (2020) 013 [arXiv:1911.12333] [INSPIRE].

[6] G. Penington, S.H. Shenker, D. Stanford and Z. Yang, *Replica wormholes and the black hole interior*, arXiv:1911.11977 [INSPIRE].

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

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

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

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

[11] A. Strominger, *Les houches lectures on black holes*, in the proceedings of the *NATO Advanced Study Institute: Les Houches Summer School, Session 62: Fluctuating Geometries in Statistical Mechanics and Field Theory*, August 2–September 9, Les Houches, France (1994), hep-th/9501071 [INSPIRE].

[12] C.G. Callan Jr., S.B. Giddings, J.A. Harvey and A. Strominger, *Evanescent black holes*, *Phys. Rev. D* **45** (1992) 1005 [hep-th/9111056] [INSPIRE].

[13] A. Bilal and C.G. Callan Jr., *Liouville models of black hole evaporation*, *Nucl. Phys. B* **394** (1993) 73 [hep-th/9205089] [INSPIRE].

[14] J.G. Russo, L. Susskind and L. Thorlacius, *The endpoint of Hawking radiation*, *Phys. Rev. D* **46** (1992) 3444 [hep-th/9206070] [INSPIRE].

[15] S.B. Giddings and A. Strominger, *Quantum theories of dilaton gravity*, *Phys. Rev. D* **47** (1993) 2454 [hep-th/9207034] [INSPIRE].

[16] S.P. de Alwis, *Quantum black holes in two-dimensions*, hep-th/9207095 [INSPIRE].

[17] J.L. Karczmarek, J.M. Maldacena and A. Strominger, *Black hole non-formation in the matrix model*, *JHEP* **01** (2006) 039 [hep-th/0411174] [INSPIRE].

[18] S. Sachdev and J. Ye, *Gapless spin fluid ground state in a random, quantum Heisenberg magnet*, *Phys. Rev. Lett.* **70** (1993) 3339 [cond-mat/9212030] [INSPIRE].

[19] A. Kitaev. http://online.kitp.ucsb.edu/online/joint98/kitaev/.

[20] J.M. Maldacena and A. Strominger, *Semiclassical decay of near extremal five-branes*, *JHEP* **12** (1997) 008 [hep-th/9710014] [INSPIRE].

[21] O. Aharony, M. Berkooz, D. Kutasov and N. Seiberg, *Linear dilatons, NS five-branes and holography*, *JHEP* **10** (1998) 004 [hep-th/9808149] [INSPIRE].

[22] Y. Zhao, *A quantum circuit interpretation of evaporating black hole geometry*, arXiv:1912.00909 [INSPIRE].

[23] C. Akers, N. Engelhardt and D. Harlow, *Simple holographic models of black hole evaporation*, arXiv:1910.00972 [INSPIRE].

[24] M. Rozali et al., *Information radiation in BCFT models of black holes*, *JHEP* **05** (2020) 004 [arXiv:1910.12836] [INSPIRE].

[25] H.Z. Chen, Z. Fisher, J. Hernandez, R.C. Myers and S.-M. Ruan, *Information flow in black hole evaporation*, *JHEP* **03** (2020) 152 [arXiv:1911.03402] [INSPIRE].

[26] R. Bousso and M. Tomašević, *Unitarity from a smooth horizon?*, arXiv:1911.06305 [INSPIRE].

[27] A. Almheiri, R. Mahajan and J.E. Santos, *Entanglement islands in higher dimensions*, *SciPost Phys.* **9** (2020) 001 [arXiv:1911.09666] [INSPIRE].

[28] Y. Chen, *Pulling out the island with modular flow*, *JHEP* **03** (2020) 033 [arXiv:1912.02210] [INSPIRE].

[29] A. Laddha, S.G. Prabhu, S. Raju and P. Shrivastava, *The holographic nature of null infinity*, arXiv:2002.02448 [INSPIRE].

[30] A. Mousatov and E. Silverstein, *Recovering infalling information via string spreading*, arXiv:2002.12377 [INSPIRE].

[31] I. Kim, E. Tang and J. Preskill, *The ghost in the radiation: robust encodings of the black hole interior*, *JHEP* **06** (2020) 031 [arXiv:2003.05451] [INSPIRE].

[32] K. Saraswat and N. Afshordi, *Extracting Hawking radiation near the horizon of AdS black holes*, arXiv:2003.12676 [INSPIRE].

[33] Y. Chen, X.-L. Qi and P. Zhang, *Replica wormhole and information retrieval in the SYK model coupled to Majorana chains*, *JHEP* **06** (2020) 121 [arXiv:2003.13147] [INSPIRE].

[34] D. Marolf and H. Maxfield, *Transcending the ensemble: baby universes, spacetime wormholes and the order and disorder of black hole information*, arXiv:2002.08950 [INSPIRE].

[35] H. Verlinde, *ER = EPR revisited: on the entropy of an Einstein-Rosen bridge*, arXiv:2003.13117 [INSPIRE].

[36] C. Jana, R. Loganayagam and M. Rangamani, *Open quantum systems and Schwinger-Keldysh holograms*, arXiv:2004.02888 [INSPIRE].

[37] S.B. Giddings and G.J. Turiaci, *Wormhole calculus, replicas and entropies*, arXiv:2004.02900 [INSPIRE].

[38] H. Liu and S. Vardhan, *A dynamical mechanism for the Page curve from quantum chaos*, arXiv:2002.05734 [INSPIRE].

[39] J. Pollack, M. Rozali, J. Sully and D. Wakeham, *Eigenstate thermalization and disorder averaging in gravity*, arXiv:2002.02971 [INSPIRE].

[40] V. Balasubramanian et al., *Geometric secret sharing in a model of Hawking radiation*, arXiv:2003.05448 [INSPIRE].

[41] F.F. Gautason, L. Schneiderbauer, W. Sybesma and L. Thorlacius, *Page curve for an evaporating black hole*, *JHEP* **05** (2020) 091 [arXiv:2004.00598] [INSPIRE].

[42] T. Anegawa and N. Iizuka, *Notes on islands in asymptotically flat 2d dilaton black holes*, arXiv:2004.01601 [INSPIRE].

[43] T.M. Fiola, J. Preskill, A. Strominger and S.P. Trivedi, *Black hole thermodynamics and information loss in two-dimensions*, *Phys. Rev. D* **50** (1994) 3987 [hep-th/9403137] [INSPIRE].

[44] G. Mandal, A.M. Sengupta and S.R. Wadia, *Classical solutions of two-dimensional string theory*, *Mod. Phys. Lett. A* **6** (1991) 1685 [INSPIRE].

[45] E. Witten, *On string theory and black holes*, *Phys. Rev. D* **44** (1991) 314 [INSPIRE].

[46] T. Piran and A. Strominger, *Numerical analysis of black hole evaporation*, *Phys. Rev. D* **48** (1993) 4729 [hep-th/9304148] [INSPIRE].

[47] A. Strominger and L. Thorlacius, *Conformally invariant boundary conditions for dilaton gravity*, *Phys. Rev. D* **50** (1994) 5177 [hep-th/9405084] [INSPIRE].

[48] J. Polchinski and A. Strominger, *A possible resolution of the black hole information puzzle*, *Phys. Rev. D* **50** (1994) 7403 [hep-th/9407008] [INSPIRE].

[49] P. Calabrese, J. Cardy and E. Tonni, *Entanglement entropy of two disjoint intervals in conformal field theory*, *J. Stat. Mech.* **0911** (2009) P11001 [arXiv:0905.2069] [INSPIRE].

[50] P. Calabrese and J. Cardy, *Entanglement entropy and conformal field theory*, *J. Phys. A* **42** (2009) 504005 [arXiv:0905.4013] [INSPIRE].

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

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

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

[54] Y. Sekino and L. Susskind, *Fast scramblers*, *JHEP* **10** (2008) 065 [arXiv:0808.2096] [INSPIRE].

[55] S.H. Shenker and D. Stanford, *Black holes and the butterfly effect*, *JHEP* **03** (2014) 067 [`arXiv:1306.0622`] [INSPIRE].

[56] A. Almheiri, R. Mahajan and J. Maldacena, *Islands outside the horizon*, `arXiv:1910.11077` [INSPIRE].

[57] A. Strominger and S.P. Trivedi, *Information consumption by Reissner-Nordstrom black holes*, *Phys. Rev. D* **48** (1993) 5778 [`hep-th/9302080`] [INSPIRE].

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

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

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

[61] C. Holzhey, F. Larsen and F. Wilczek, *Geometric and renormalized entropy in conformal field theory*, *Nucl. Phys. B* **424** (1994) 443 [`hep-th/9403108`] [INSPIRE].

[62] P. Calabrese and J.L. Cardy, *Entanglement entropy and quantum field theory*, *J. Stat. Mech.* **0406** (2004) P06002 [`hep-th/0405152`] [INSPIRE].

[63] R.C. Myers, *Black hole entropy in two-dimensions*, *Phys. Rev. D* **50** (1994) 6412 [`hep-th/9405162`] [INSPIRE].