Published for SISSA by Springer

Received: *March 13, 2023*
Revised: *June 22, 2023*
Accepted: *June 22, 2023*
Published: *July 5, 2023*

The holographic map of an evaporating black hole

Zsolt Gyongyosi,[a] **Timothy J. Hollowood,**[a] **S. Prem Kumar,**[a] **Andrea Legramandi**[a,b,c] **and Neil Talwar**[a]

[a] *Department of Physics, Swansea University,*
Swansea, SA2 8PP, U.K.

[b] *Pitaevskii BEC Center, CNR-INO and Dipartimento di Fisica, Università di Trento,*
I-38123 Trento, Italy

[c] *INFN-TIFPA, Trento Institute for Fundamental Physics and Applications,*
Trento, Italy

E-mail: z.gyongyosi.2133547@swansea.ac.uk, t.hollowood@swansea.ac.uk, s.p.kumar@swansea.ac.uk, andrea.legramandi@unitn.it, n.talwar.2017429@swansea.ac.uk

Abstract: We construct a holographic map that takes the semi-classical state of an evaporating black hole and its Hawking radiation to a microscopic model that reflects the scrambling dynamics of the black hole. The microscopic model is given by a nested sequence of random unitaries, each one implementing a scrambling time step of the black hole evolution. Differently from other models, energy conservation and the thermal nature of the Hawking radiation are taken into account. We show that the QES formula follows for the entropy of multiple subsets of the radiation and black hole. We further show that a version of entanglement wedge reconstruction can be proved by computing suitable trace norms and quantum fidelities involving the action of a unitary on a subset of Hawking partners. If the Hawking partner is in an island, its unitary can be reconstructed by a unitary on the radiation. We also adopt a similar setup and analyse reconstruction of unitaries acting on an infalling system.

Keywords: AdS-CFT Correspondence, Black Holes, Models of Quantum Gravity, 2D Gravity

ArXiv ePrint: 2301.08362

https://doi.org/10.1007/JHEP07(2023)043

Contents

1 Introduction

Black holes lie at the front line of the struggle to unify quantum mechanics with gravity. Recent progress is focused on how this struggle plays out at the level of effective theory in a gravitating system like a black hole. In particular, the effective description involves techniques that have evolved over many years involving quantum field theory over a fixed background spacetime using semi-classical techniques. In a black hole geometry this leads to the emission of Hawking radiation and the apparent loss of unitarity [1, 2]. On the other hand, there is a microscopic level of description, for example provided by string theory, in which a black hole is described as a quantum system with a large density of states given by the Bekenstein-Hawking (BH) entropy (see [3] for a review).

1.1 The holographic map

Recent progress has shed light on how these two levels of description are related and how the information-loss paradox is resolved and unitarity is restored [4–7] (also see the reviews [8, 9]). A key ingredient is a map, the 'holographic map', between the effective semi-classical description and the microscopic description

$$V: \quad \mathcal{H}^{\text{sc}} \to \mathcal{H}^{\text{micro}} \,. \tag{1.1}$$

The idea of such a map between the semi-classical and microscopic descriptions naturally arises in holography where the semi-classical state describes the state of bulk gravitational theory while the microscopic state describes the non-gravitational CFT dual [10–15]. It is becoming clear that such a map should apply more generally and specifically in spacetimes which are not asymptotically AdS, such as an evaporating black hole, where the radiation can escape the AdS bulk. The holographic map has been interpreted as the encoding map of a quantum error code and this synergy between the two subjects has been very fruitful and has led to a better understanding of entanglement wedge reconstruction [16–26]. However, recent work [27, 28][1] has clarified certain details and in particular argued that, in the context of a black hole, it is an important feature that the map is not isometric, $V^\dagger V \neq 1$. This means that the relation with the standard theory of quantum error correcting codes is not so compelling. The non-isometric nature of the map is actually very natural because as the black hole ages its Hilbert space becomes too small to accommodate all the Hawking partners of the previously emitted radiation and so something has to give. Another key insight of [28] is that the map does not act on the radiation once it has dispersed away from the black hole. This clarifies certain statements that have been made about the radiation, in particular it is not possible to change the microscopic state of the black hole by making operations on the radiation however complicated: there is no long-range non-locality of this kind.

The purpose of this work is to construct the holographic map V in a very simple microscopic model of black hole evaporation defined e.g. in [31, 32] but refined to take account of energy conservation leading to thermal states. The basic version of the model is the block random unitary model (BRU) of [28]. A number of key features follow also for this more refined model:

① The semi-classical state of the radiation ρ_R^{sc} is precisely the average of the microscopic state of the radiation ρ_R over the quasi-random microscopic scrambling dynamics of the black hole.

② Past the Page time the quasi-random fluctuations of the microscopic state ρ_R overwhelm the state and it becomes very different from the semi-classical (Hawking) state ρ_R^{sc}.

③ The Quantum Extremal Surface (QES) formula [33, 34] for the entropy of a generic number of radiation and black hole subsets is derived in the regime where the black hole is evaporating slowly [31, 35].

[1] See also [29, 30] for recent developments.

④ Unitary actions on an infalling system can be reconstructed on the radiation after the Page time showing that the information of the infalling system has been teleported out of the black hole realizing the Hayden-Preskill 'black hole as a mirror' scenario [36].

⑤ There is a version of state-specific entanglement wedge reconstruction (of the type discussed in [28]): local unitaries acting on the Hawking partners can be reconstructed as a unitary acting on the black hole before the Page time and on the radiation after the Page time. The discussion is extended for generic subsets of the Hawking radiation.

Let us now put some flesh on the bones. At the semi-classical level, the state of a QFT in the black hole background consists of an entangled state between the outgoing Hawking radiation R and their partner modes behind the horizon $\overline{R}$. The overall state is pure

$$|\psi\rangle = \Big\{\sum_J \lambda_J |J\rangle_R \otimes |J\rangle_{\overline{R}}\Big\} \otimes |S\rangle_F \in \mathcal{H}^{\text{sc}}\,. \tag{1.2}$$

We have also included the possibility for infalling modes in the state $|S\rangle_F$, including the matter that collapsed to form the black hole. We will develop two models: (i) a simple one in which the Hilbert space of the radiation is taken to be finite dimensional and (1.2) is the maximally entangled state $\lambda_J = 1/\sqrt{d_R}$ and (ii) a more refined one for which the radiation and partners are in a thermofield double with a slowly varying temperature.

At the microscopic level, the black hole is described by a finite dimensional Hilbert space $\mathcal{H}_B$ whose dimension is exponential in the BH entropy $d_B = e^{S_{\text{BH}}}$. The black hole emits Hawking radiation and at the microscopic level we can write the state of a partly evaporated black hole and radiation as

$$|\Psi\rangle = \sum_J \lambda_J |J\rangle_R \otimes |\Psi_J\rangle_B \in \mathcal{H}^{\text{micro}}\,. \tag{1.3}$$

The two states, the semi-classical $|\psi\rangle$ and the microscopic $|\Psi\rangle$ are related by the holographic map (1.1)

$$V: \quad \mathcal{H}_R \otimes \mathcal{H}_{\overline{R}} \otimes \mathcal{H}_F \to \mathcal{H}_R \otimes \mathcal{H}_B\,. \tag{1.4}$$

It was argued in [28] that the map should act trivially on R because the outgoing radiation system is identical in both the semi-classical and microscopic descriptions. So V actually only acts non-trivially as $\mathcal{H}_{\overline{R}} \otimes \mathcal{H}_F \to \mathcal{H}_B$. This is natural because the Hawking partner modes $\overline{R}$ and the infalling modes F are behind the horizon and so part of the black hole whose semi-classical geometry should emerge from the microscopic description. By comparing (1.2) with (1.3), we have

$$V|J\rangle_{\overline{R}} \otimes |S\rangle_F = |\Psi_J\rangle_B\,. \tag{1.5}$$

We leave the dependence on the infalling state implicit.

The way that Hawking's information loss paradox can be resolved now reveals itself. In Hawking's analysis, the state of the radiation is the reduced state, the maximally-mixed state in the basic model and a quasi-thermal state in the refined model

$$\rho_R^{\text{sc}} = \sum_J |\lambda_J|^2 |J\rangle_R\langle J|\,, \tag{1.6}$$

since the partner mode states are orthonormal, ${}_{\overline{R}}\langle J|K\rangle_{\overline{R}}=\delta_{JK}$. On the other hand, at the microscopic level,

$$\rho_R=\sum_{JK}\lambda_K\bar{\lambda}_J\xi_{KJ}|K\rangle_R\langle J|\,,\qquad \xi_{KJ}=\langle\Psi_J|\Psi_K\rangle\,. \tag{1.7}$$

The semi-classical state is devoid of internal correlations, information is lost and unitarity is violated. The microscopic state, on the other hand, can carry the correlations and repair unitarity if the inner products ξ_{JK} are non-trivial. The fact that

$$\langle\Psi_J|\Psi_K\rangle\neq\delta_{JK}\,, \tag{1.8}$$

implies that the holographic map V is non-isometric, a key insight in [28]. It is the non-isometric nature of V that allows information to escape out of the black hole in the correlations induced by the inner product [37]. Such a release of information would presumably be interpreted as being a non-local process to a semi-classical observer. This is a major insight but perhaps to be expected when spacetime geometry is an emergent concept.

For a black hole past its Page time, when $S_\text{rad}\gg S_\text{BH}$, one would expect the states $|\Psi_J\rangle$ to be far from orthogonal because there are order $e^{S_\text{rad}(R)}$ states in a much smaller e^{S_BH} dimensional Hilbert space. Roughly speaking, as previously argued e.g. in [4, 38], we find

$$\langle\Psi_J|\Psi_K\rangle=\begin{cases}1+\mathscr{O}(e^{-S_\text{BH}}) & J=K\,,\\ \mathscr{O}(e^{-S_\text{BH}/2}) & J\neq K\,,\end{cases} \tag{1.9}$$

so the violation appears to be exponentially small $\sim e^{-1/G}$ in the semi-classical limit. This seems to suggest that the corrections coming from the microscopic theory will be small. However, if we write $\xi=I+Z$, then Z is roughly-speaking a quasi-random Hermitian matrix whose elements are order $e^{-S_\text{BH}/2}$. It seems, therefore, that the effect of Z would be very suppressed. However, if the dimension of the matrix $\sim e^{S_\text{rad}}$ is large then its eigenvalues can be expected to lie in a distribution between $\pm e^{(S_\text{rad}-S_\text{BH})/2}$. What this indicates is that the fluctuations in Z could be expected to give rise to a radical change in the state of the radiation beyond the Page time when $S_\text{rad}\gg S_\text{BH}$ and a mechanism to ensure the unitarity of the evaporation. On the other hand, if we average the microscopic state over the quasi-random fluctuations Z we recover the semi-classical state

$$\overline{\rho_R}=\rho_R^\text{sc}\,. \tag{1.10}$$

The fact that $\rho_R\neq\rho_R^\text{sc}$ means that if we were to attempt to interpret the microscopic state as a state on the semi-classical geometry, then in the near-horizon region it would not be the inertial vacuum and so we could expect there will be non-trivial energy and momentum as the horizon is approached [39].

Another issue that is clarified by the fact that V acts trivially on the radiation R is, as already mentioned, that the state of black hole is completely invariant under any local action on the radiation. In more detail, the most general local action is obtained by coupling R to an auxiliary system M and having them interact. On the semi-classical state

$$|\psi\rangle\otimes|\varnothing\rangle_M\longrightarrow\sum_\alpha K_\alpha|\psi\rangle\otimes|\alpha\rangle_M\,, \tag{1.11}$$

for some orthonormal states $|\alpha\rangle$ of M and where the operators K_α act on R. This defines a quantum channel acting on R and unitarity implies that K_α are Krauss operators $\sum_\alpha K_\alpha^\dagger K_\alpha = 1$. Mapping this to the microscopic state, and using the fact that $[V, K_\alpha] = 0$, the reduced state on B, after R and M have interacted, is

$$\rho'_B = \sum_\alpha \text{Tr}_R\Big\{K_\alpha|\Psi\rangle\langle\Psi|K_\alpha^\dagger\Big\} = \text{Tr}_R\left\{|\Psi\rangle\langle\Psi|\sum_\alpha K_\alpha^\dagger K_\alpha\right\} = \rho_B\,, \tag{1.12}$$

so the state of the black hole is invariant.

1.2 The QES formula

One can quantitatively appreciate how ρ_R differs from ρ_R^{sc} by calculating their von Neumann entropies. The entropy of the semi-classical state ρ_R^{sc}, suitably regularized, is just the thermal entropy of Hawking radiation familiar from Hawking's calculation. The question is, how to calculate the entropy of the microscopic state ρ_R? This is where the QES, or generalized entropy, formula comes in [14, 33, 40–42]. It relates the von Neumann entropy of the microscopic state ρ_A reduced on some subsystem factor e.g. $A = R$ or B, or some more specific subset of R to the generalised entropy:

$$S(\rho_A) = \min S_{\text{gen}}(X_A), \qquad S_{\text{gen}}(X_A) = \frac{\mathscr{A}(X_A)}{4G} + S(\rho^{\text{sc}}_{\mathcal{W}(A)}). \tag{1.13}$$

Here $\mathscr{A}(X_A)$ is the area of a codimension two surface X_A in the gravitating region, called the Quantum Extremal Surface (QES), that bounds a region known as the entanglement wedge $\mathcal{W}(A)$ of A in the gravitating region. The full entanglement wedge $\mathcal{W}(A)$ is given by appending $A \cap R$ to this region[2] and is determined by extremising the generalised entropy. Note that if A is the radiation R, or some subset thereof, the entanglement wedge $\mathcal{W}(A)$ consists of A and potentially also a region disconnected from A i.e. $\mathcal{W}(A) = A \cup I$. The region I is known as the 'entanglement island', or 'island' for short.

The formula (1.13) is remarkable in several ways but principally because it allows one to calculate the entropy of the microscopic state ρ_A using only semi-classical techniques even when the details of the microscopic theory are not known. It does this by implicitly averaging over the complex chaotic microscopic dynamics of the black hole in the way familiar from statistical mechanics. More precisely, when computed in the semi-classical theory, we can think of the left hand side as being equal to the usual $n \to 1$ limit of the Rényi entropies but averaged in the following way

$$S(\rho_A) = \lim_{n\to 1}\frac{1}{1-n}\log \overline{e^{(1-n)S^{(n)}(\rho_A)}}\,, \tag{1.14}$$

with the average over a suitable ensemble that is a proxy for the underlying complex, chaotic microscopic dynamics. Just as in statistical mechanics, the conceptual idea is that the average captures the behaviour of a single typical microscopic state because, unlike the state itself, the Rényi entropies are self-averaging quantities.

[2] More precisely, $\mathcal{W}(A)$ is the domain of dependence of this region.

For an evaporating black hole, the QES are behind the horizon and when the evaporation is slow, which it is for most of the evaporation time apart from the final stage, the QES are very close behind the horizon. In fact the QES are completely determined within the scope of the slow evaporation approximation [31, 35, 43]. Firstly, they have Kruskal-Szekeres (KS) coordinates related via

$$UV \sim \frac{c}{S_{\text{BH}}} \ll 1\,, \tag{1.15}$$

where c are the number of (massless) fields. In terms of Eddington-Finkelstein (EF) coordinates (u, v),[3] this means

$$v = u - \Delta t_{\text{scr}}\,, \qquad \Delta t_{\text{scr}} = \frac{1}{2\pi T}\log\frac{S_{\text{BH}}}{c}\,. \tag{1.16}$$

The slow evaporation regime applies precisely when $S_{\text{BH}} \gg c$ so that the QES are pressed up against the horizon from within. The time shift above between the infalling and outgoing coordinates Δt_{scr} is identified with the scrambling time of the black hole. This is time dependent but only changes slowly as the black hole evaporates.

The second condition on the QES is that the outgoing EF coordinate of a QES u_{QES} (inside the horizon) must be equal to the outgoing EF coordinate of one of the endpoints of the radiation $u_{\partial A}$ (outside the horizon)

$$\{u_{\text{QES}}\} \subset \{u_{\partial A}\}\,. \tag{1.17}$$

This reduces the variation problem to a discrete minimization problem.

When A is a subset of the radiation and the entanglement wedge $\mathcal{W}(A) = A \cup I$, the second term in (1.13) is just the thermal entropy[4]

$$S(\rho^{\text{sc}}_{\mathcal{W}(A)}) \approx S_{\text{rad}}(A \ominus \tilde{I}) = \frac{\pi c}{6}\int_{A\ominus\tilde{I}} T(u)\,du\,, \tag{1.18}$$

where $T(u)$ is the instantaneous temperature of the black hole as a function of the outgoing EF coordinate u on $\mathscr{I}^+$. Here, $\tilde{I}$, the 'island-in-the stream', is just the reflection of the island in the horizon and projected onto $\mathscr{I}^+$ [31, 35, 44]. So in terms of the outgoing EF coordinate u, I and $\tilde{I}$ are equal, with the former outside the horizon and the latter inside. The symmetric difference in (1.18) accounts for the fact that I contains purifiers of the radiation. The first term in (1.13) is then approximately equal to the Bekenstein-Hawking entropy S_{BH} evaluated at EF outgoing coordinates of the QES $u_{\partial I}$. Hence, within the slow evaporation approximation, we can write the entropy as a discrete minimization problem

$$S\,(A) \approx \min_I \left\{\sum_{u_{\partial I}} S_{\text{BH}}(u_{\partial I}) + S_{\text{rad}}(A \ominus \tilde{I})\right\}\,. \tag{1.19}$$

[3] The KS and EF coordinates are related by an approximately exponential map, $U = -\exp\left(-2\pi\int^u T(t)dt\right)$ and $V = \exp\left(2\pi\int^v T(t)dt\right)$, where $T(t)$ is the instantaneous temperature of the black hole.

[4] There is a common divergence associated with the end-points of A at $\mathscr{I}^+$ which can be regularized. The divergences associated to end-points of I, on the other hand, are precisely cancelled by the divergences in the area term in (1.13).

This formula can easily be adapted to the case when A includes the black hole itself, $B \subset A$. One simply replaces A by $A \cap R$ in the second term.[5] In section 3 we verify this formula in both the basic and refined models using the replica trick. We also find a simple formula (3.29) for the island which relates the replica trick and the entanglement wedge in quite a direct way.

The paper is organized as follows. In section 2 we define two simple discrete models of a holographic map for an evaporating black hole. There is a basic and refined model. Compared with the basic model, the refined model has the nice features that the state of a small subsystem is thermal instead of maximally mixed and that the irreversiblity of evaporation is naturally incorporated (there is no need to add in ancilla qubits to mimic this effect). We also show that the average over the quasi-random unitary time evolution of the microscopic state is just the semi-classical state (1.10). In section 3, we compute the Rényi and von Neumann entropies of subsets of the radiation and black hole and derive the minimization problem for the generalized entropy of a slowly evaporating black hole (1.19). In section 4, we turn to the Hayden-Preskill scenario [36] and consider when the action of a unitary on an infalling system be reconstructed (in a state-specific sense) on a subset of the radiation or black hole from a 'decoupling argument'. We find the model reproduces the 'black hole as a mirror' phenomenon and reconstruction is possible on the radiation when the black hole is past the Page time. This problem of reconstruction of operators acting on an infalling system was studied in the basic (or BRU) model and a random pairwise interaction model (which incorporates the fast scrambling nature of black holes) in [28]. Our main contributions here are to study this problem in a model which generalises the basic model and also to consider when reconstruction is possible not just on the radiation or the black hole, but a subset thereof. In section 5 we consider when local operations, in the form of a quantum channel, acting on the Hawking partners can be reconstructed on a subset of the radiation or black hole. As expected, we find that reconstruction on the radiation is possible when the black hole is past the Page time. In section 6 we draw some conclusions. In appendix A we review the computation of certain thermodynamic quantities for free bosonic and fermionic fields, which is used in the refined model. In appendix B we provide a proof of the dominant saddles which contribute in the replica trick calculation in the basic model.

2 The model

In the model, described in [31], the evaporation at the microscopic level is described by a series of discrete time steps identified with the scrambling time of the black hole (1.16) shown in the figure 1. During the p^{th} time step the state of the black hole evolves by a unitary U_p which maps

$$U_p : \quad \mathcal{H}_{B_{p-1}} \otimes \mathcal{H}_{F_{p-1}} \longrightarrow \mathcal{H}_{B_p} \otimes \mathcal{H}_{R_p} . \tag{2.1}$$

[5] Then, if $R_N \notin A$, the most recent emitted interval of radiation, there must be a QES with a u coordinate equal to the u coordinate of the upper end-point of R_N, giving a contribution $S_{\text{BH}}(M_N) \equiv S_{\text{BH}}(M)$ to (1.19). On the other hand, if $R_N \in A$ then it must be that $R_N \subset I$. In the latter case, the connected subset of I that includes R_N is not strictly-speaking part of the island although it is in the entanglement wedge of A.

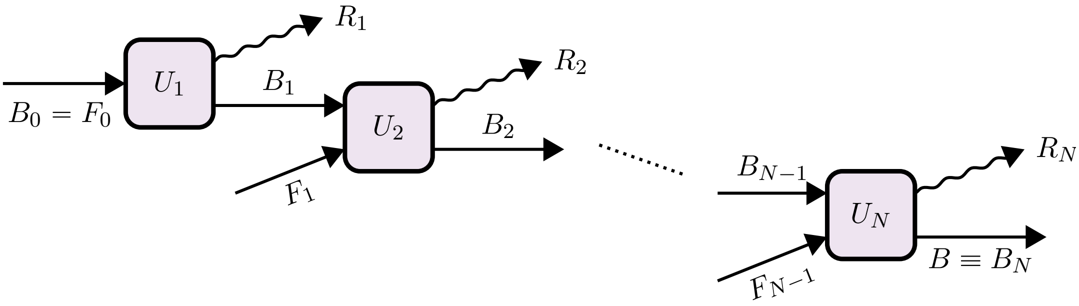

Figure 1. The model of black hole evaporation consisting of a sequence of random unitaries that mimic the scrambling microscopic dynamics. At each time step a small subsystem escapes as the Hawking radiation and there can be an infalling system. The time steps are of the order of the scrambling time of the black hole (1.16) and so the model will appear to be continuous at time scales much larger than the scrambling time, including the Page time and the evaporation time.

In the basic model, we have $d_{B_{p-1}} d_{F_{p-1}} = d_{B_p} d_{R_p}$, whereas in the refined model energy conservation is taken into account and R_p is infinite dimensional. In this case, U_p is an isometric embedding of a microcanonical energy window into $\mathcal{H}_{R_p} \otimes \mathcal{H}_{B_p}$, as we will describe later. After N time steps, the state of the black hole and radiation is

$$|\Psi(t_N)\rangle = U_N \cdots U_2 U_1 |S\rangle_F \in \mathcal{H}_B \otimes \mathcal{H}_R \,, \tag{2.2}$$

where

$$|S\rangle_F = |s_0\rangle_{F_0} \otimes |s_1\rangle_{F_1} \otimes \cdots \otimes |s_{N-1}\rangle_{F_{N-1}} \,, \tag{2.3}$$

describes the infalling matter that created the black hole $B_0 \equiv F_0$ as well as matter that falls in during each time step F_p as the black hole evaporates. The radiation is split into a temporal sequence of subsets $R = \bigcup_p R_p$. In the above, the remaining black hole is $B \equiv B_N$. A basis of states of the radiation consists of $|J\rangle_R$ where $J = \{j_1, \dots j_N\}$ and each $j_p \in \{1, 2, \dots, d_{R_p}\}$ labels the states in the p^{th} time step R_p. In particular, the microscopic states defined in (1.3) are

$$|\Psi_J\rangle = \frac{1}{\lambda_J}\, {}_R\langle J| U_N \cdots U_2 U_1 |S\rangle_F \in \mathcal{H}_B \,. \tag{2.4}$$

Consequently the time evolution of the black hole in the model leads to a concrete expression for the holographic map,

$$V = \sum_J \frac{1}{\lambda_J}\, {}_{\overline{R}}\langle J| \otimes {}_R\langle J| U_N \cdots U_2 U_1 \,, \tag{2.5}$$

acting on $\mathcal{H}_{\overline{R}} \otimes \mathcal{H}_F$. In the basic model we take $\lambda_J = 1/\sqrt{d_R}$ and so the map has the form of a unitary followed by a post selection on the maximally-entangled state on $\mathcal{H}_{\overline{R}} \otimes \mathcal{H}_R$. In the refined model we take λ_J to be given by (2.15). In the refined model the map also has the form of a unitary followed by a post selection, however, we note that the post selection is not on the thermofield double state on $\mathcal{H}_{\overline{R}} \otimes \mathcal{H}_R$. In fact, in the refined model the sum in (2.5) is not well defined and V is only defined acting on suitable states such as $|\psi\rangle$.

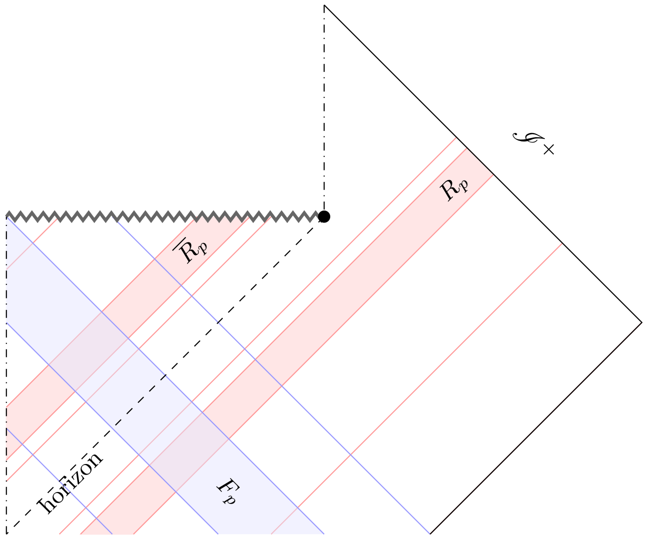

Figure 2. Subsets of Hawking modes R_p are their entangled partners $\overline{R}_p$ behind the horizon and infalling modes F_p. The Hawking modes propagate out to null infinity $\mathscr{I}^+$. Each R_p and F_p lasts for a scrambling time.

The basic model [31] is identified with the block random unitary model of [28]. What is noteworthy is that, in the basic model, the post selection on the maximally entangled state, which manifests the non-isometric property of the map, is also the mechanism which allows information to be teleported out of the black hole.

At the semi-classical level, the subsets of Hawking radiation R_p and their partners behind the horizon $\overline{R}_p$ are illustrated in the Penrose diagram figure 2.

2.1 The refined model

In this section we refine our model of black hole evaporation to take account of energy conservation and the thermal nature of the Hawking radiation. We will work in the adiabatic, or quasi-static, regime where the black hole is evaporating slowly enough that it makes sense to ascribe a slowly varying temperature $T(t)$ to the Hawking radiation determined by the thermodynamic equation of the black hole

$$\frac{1}{T} = \frac{dS_{\text{BH}}}{dM}\,, \tag{2.6}$$

where M is the black hole mass.[6] The adiabatic regime is where Hawking's calculation derivation is valid. It is defined by the requirement that

$$S_{\text{BH}} \gg c\,, \tag{2.7}$$

the number of massless fields. In addition, for a semi-classical limit $c \gg 1$.

[6] For a Schwarzschild black hole, M is the mass, while for the charged black hole and the black hole in JT gravity, M is the mass minus the mass of the extremal black hole.

The time dependence of the energy is determined by the energy flux of the Hawking radiation. Since most of the energy loss occurs in the s-wave modes we have effectively a 1 + 1-dimensional relativistic gas. We also ignore the possibility for back-scattering of modes and so take a trivial greybody factor. The energy balance equation is then

$$\frac{dM}{dt} = -\frac{\pi c T^2}{12} \tag{2.8}$$

and given (2.6) all that it needed to determine the time evolution of M, T and S_{BH} is the energy dependence of the BH entropy which depends on the nature of the black hole. For example, for Schwarzschild $S_{\text{BH}} = 4\pi G M^2$.

We will model the evaporation in terms as a series of time steps whose size are of the order of the scrambling time of the black hole,

$$\Delta t \sim \frac{1}{T} \log \frac{S_{\text{BH}}}{c} \,. \tag{2.9}$$

Note that this is time dependent, so the size of the time steps adapt as the evaporation proceeds.

At each time step, the radiation carries away a small amount of energy in a distribution that is strongly peaked around an average. Therefore, we can model the state of the black hole at each time step as lying in a Hilbert space $\mathcal{H}_{B_p}$ describing a system with energy in a small window $\Theta_p = [M_p, M_p + \delta M]$. Implicitly, M_p includes the energy of the infalling system F_p. In other words, the black hole is in a microcanonical state. The size of the window δM is assumed to be small but for simplicity we will assume that it is much larger than the spread of the energy carried away by the radiation at each time step. The fact that the BH entropy is so large means that Θ_p contains a vast number of states that forms a quasi-continuum. The dimension of this space is exponential in the Bekenstein-Hawking entropy

$$d_{B_p} = \frac{C\delta M}{M_p} e^{S_{\text{BH}}(M_p)} \sim e^{S_{\text{BH}}(M_p)} \,. \tag{2.10}$$

In the above, C is some constant which we do not have to specify since $S_{\text{BH}}(M)$ is very large.

The picture of the black hole evolving through a sequence of microcanonical states is of course an approximation which is justified because the radiation emitted during a time step has a sharply defined average energy and a spread that is assumed to be much smaller than the width of the windows δM. Let us justify this claim. Since the time step, the scrambling time Δt, is much greater than the thermal scale T^{-1}, the energy and entropy of the Hawking radiation follow from the standard statistical mechanics of a relativistic bosonic or fermionic gas (summarized in appendix A). For a bosonic gas

$$\mathcal{E} = c\mathscr{V} \int \frac{d\omega}{2\pi} \frac{\omega}{e^{\omega/T} - 1} = \frac{\pi c \mathscr{V} T^2}{12} \tag{2.11}$$

and the entropy $S_{\text{rad}} = \pi c \mathscr{V} T/6$, where we identify the volume with the space filled by the gas in the scrambling time, i.e. $\mathscr{V} = \Delta t$. In particular, the entropy

$$S_{\text{rad}} = \frac{\pi c \Delta t T}{6} \sim c \log \frac{S_{\text{BH}}}{c} \gg 1 \,. \tag{2.12}$$

Hence, the Hawking modes emitted in a time step have a large entropy and so can be described thermodynamically. Indeed, the normalized spread of the energy

$$\frac{\Delta\mathcal{E}}{\mathcal{E}} \sim \frac{1}{\sqrt{S_{\rm rad}}} \ll 1\,. \tag{2.13}$$

On the other hand, the radiation is a much smaller system than the black hole because

$$S_{\rm BH} \gg c\log\frac{S_{\rm BH}}{c}\,. \tag{2.14}$$

We will then assume that this spread is much smaller than the microcanonical energy window $\delta M \gg \Delta\mathcal{E}$ justifying the evaporation as a sequence of microcanonical states.

The semi-classical state is now a thermofield double with a slowly varying temperature. Taking the basis states $|j_p\rangle$ to be approximate energy eigenstates with eigenvalues $\mathcal{E}_{j_p}$, we have

$$\lambda_J = \frac{e^{-\sum_p \mathcal{E}_{jp}/2T_p}}{\sqrt{\mathcal{Z}}}\,, \tag{2.15}$$

where $\mathcal{Z} = \sum_J e^{-\sum_p \mathcal{E}_{jp}/T_p}$ is the partition function which provides normalization. The temperature T_p is the instantaneous temperature of the Hawking radiation given in (2.6) evaluated at $E = M_p$. The states $|j_p\rangle$ are to be thought of as localized in an outgoing shell of thickness Δt_p. This is justified because the modes have characteristic momentum T_p and so can be localized on scales T_p^{-1} which is much smaller than Δt_p.

2.2 The average state

Black holes are famously fast scramblers so that over the scrambling time U_p is essentially a random unitary. The question of how random time evolution of a black hole is an interesting question but one can make the hypothesis that for certain quantities it is effectively indistinguishable from a Haar random unitary. In this section, we make that assumption and compute the average of the microscopic state. We note that this was analysed in the basic model in [28] and we review the calculation here to set up some notation.

We will need to average quantities over an $\mathcal{N}\times\mathcal{N}$ unitary for which the basic results is the integral

$$\int dU\; U^*_{AB}U_{A'B'} = \frac{1}{\mathcal{N}}\delta_{AA'}\delta_{BB'}\,. \tag{2.16}$$

We will also need the generalization of this involving n replicas:

$$\int dU\;\prod_{j=1}^n U^*_{A_jB_j}U_{A'_jB'_j} = \sum_{\sigma,\tau\in S_n}\prod_{j=1}^n \delta_{A_jA_{\sigma(j)}}\delta_{B_jB_{\tau(j)}}Wg(\sigma\tau^{-1},\mathcal{N})\,, \tag{2.17}$$

where Wg is the Weingarten function [55, 56]. Note how the integrals over the replicas involves a sum over the elements of the symmetric group $\sigma,\tau\in S_n$ that permute the replicas. We will only need the behaviour in the limit that $\mathcal{N}$ is large, which picks out the terms with $\sigma=\tau$ for which $Wg(1,\mathcal{N}) = 1/\mathcal{N}$,

$$\int dU\;\prod_{j=1}^n U^*_{A_jB_j}U_{A'_jB'_j} = \frac{1}{\mathcal{N}^n}\sum_{\tau\in S_n}\prod_{j=1}^n \delta_{A_jA'_{\tau(j)}}\delta_{B_jB'_{\tau(j)}} + \cdots\,, \tag{2.18}$$

Let us consider the microscopic state of the radiation ρ_R and compute its average over the unitaries U_p, $p = 1, 2, \ldots, N$. The ket $|\Psi\rangle$ contributes a U_p and bra $\langle\Psi|$ a $U_p^\dagger$. The average over U_p then knits together the bra and ket.

Let us focus on the average over U_p of its adjoint action on a operator f. Using (2.16), we can write this average as

$$\int dU_p \, U_p f U_p^\dagger = \mathrm{Tr}(f)\rho^{(\mathrm{mm})}_{R_pB_p} \, . \tag{2.19}$$

Here, $\rho^{(\mathrm{mm})}_{R_pB_p}$ is the maximally-mixed state on $\mathcal{H}_{R_p} \otimes \mathcal{H}_{B_p}$ which in the basic model is,

$$\rho^{(\mathrm{mm})}_{R_pB_p} = \frac{\mathbf{1}}{d_{R_p}d_{B_p}} \, . \tag{2.20}$$

In the refined model it is the maximally-mixed state in the energy window Θ_{p-1} embedded in $\mathcal{H}_{R_p} \otimes \mathcal{H}_{B_p}$ in such a way as to conserve energy,

$$\rho^{(\mathrm{mm})}_{R_pB_p} \propto \Pi_{\Theta_{p-1}} \, . \tag{2.21}$$

where $\Pi_{\Theta_{p-1}}$ is the projector onto the energy window. The following U_{p+1} average then imposes a trace over B_p. In the basic model, that gives

$$\mathrm{Tr}_{B_p}\big(\rho^{(\mathrm{mm})}_{R_pB_p}\big) = \frac{1}{d_{R_p}} \sum_{j_p} |j_p\rangle\langle j_p| \, , \tag{2.22}$$

In the refined model, let us denote a basis of energy eigenstates of R_p as $|j_p\rangle$ with energies $\mathcal{E}_{j_p}$, then

$$\mathrm{Tr}_{B_p}\big(\rho^{(\mathrm{mm})}_{R_pB_p}\big) = e^{-S_{\mathrm{BH}}(M_{p-1})} \sum_{j_p} e^{S_{\mathrm{BH}}(M_{p-1}-\mathcal{E}_{j_p})} |j_p\rangle\langle j_p| \, . \tag{2.23}$$

Implicitly, the sum here is constrained to have $M_{p-1} - \mathcal{E}_{j_p} \in \Theta_p$. We can now follow the standard route for deriving the canonical ensemble of a small subsystem of a larger system in a microcanonical state [45], in our case the maximally mixed state. Since the radiation subsystem is much smaller then the black hole, we can expand $S_{\mathrm{BH}}(M_{p-1} - \mathcal{E}_{j_p}) \approx S_{\mathrm{BH}}(M_{p-1}) - \mathcal{E}_{j_p}/T_p$ where the temperature is defined in the standard way via the thermodynamic equation (2.6) for a black hole of mass M_{p-1}. Then we can extend the restricted sum over $\mathcal{E}_{j_p}$ to be unrestricted because terms for which $M_{p-1} - \mathcal{E}_{j_p} \notin \Theta_p$ are heavily suppressed. This gives the familiar approximation, namely the canonical state

$$\mathrm{Tr}_{B_p}\big(\rho^{(\mathrm{mm})}_{R_pB_p}\big) \approx \sum_{j_p} \frac{e^{-\mathcal{E}_{j_p}/T_p}}{\mathcal{Z}_p} |j_p\rangle\langle j_p| \, , \tag{2.24}$$

where $\mathcal{Z}_p = \sum_{j_p} e^{-\mathcal{E}_{j_p}/T_p}$.

If we now assemble the expressions (2.22) and (2.24) for all the time steps, to find the average state of the radiation

$$\begin{aligned} \overline{\rho_R} &= \frac{1}{d_R}\sum_J |J\rangle\langle J| && \text{(basic)}\,, \\ \overline{\rho_R} &= \sum_J \frac{e^{-\sum_p \mathcal{E}_{jp}/T_p}}{\mathcal{Z}} |J\rangle\langle J| && \text{(refined)}\,. \end{aligned} \tag{2.25}$$

Hence the averaged microscopic state ρ_R is precisely the semi-classical state ρ_R^{sc} as stated in (1.10). Deviations from the average arise because of the non-isometric nature of the map. These have been analysed in the basic model in [28].

3 Entropies

We can calculate the entropy of the microscopic state reduced on any subset

$$A \subset \{R_1, \dots, R_N, B\}\,. \tag{3.1}$$

The strategy is to first calculate the Rényi entropies which can be defined by introducing n replicas of the Hilbert space

$$e^{(1-n)S^{(n)}(A)} = \text{Tr}(\rho_A^n) = \text{Tr}^{(n)}\sigma_1^{[R_1]}\cdots\sigma_N^{[R_N]}\tau_{N+1}^{[B]}\,|\Psi\rangle\langle\Psi|^{\otimes n}\,, \tag{3.2}$$

where the σ_p and τ_{N+1} are elements of the symmetric group S_n. The superscripts, e.g. $\sigma_p^{[R_p]}$, on these elements indicate which subspace of the replicated Hilbert space the element acts on where it is ambiguous. These elements are taken to be either the identity element e or the cyclic permutation η according to the definition of the subset A

$$A = \big\{R_p \,\big|\, \sigma_p = \eta\,,\; p = 1, 2, \dots, N\big\} \cup \big\{B \,\big|\, \tau_{N+1} = \eta\big\}\,. \tag{3.3}$$

The Rényi entropies are known to be self-averaging in the ensemble of the unitaries U_p (e.g. [46]) and so we will calculate the ensemble average of (3.3) and take this to describe a typical element of the ensemble. The integrals we need are given in (2.18) which capture the leading order behaviour when the Hilbert spaces have a large dimension.

Using (2.18), the average over the unitary U_p acting in a replicated Hilbert space at large $d_{R_p}d_{B_p}$ of adjoint action is given by a sum over elements of the symmetric group S_n,

$$\int dU_p\, U_p^{\dagger\otimes n}\, f\, U_p^{\otimes n} = \sum_{\tau_p\in S_n}\Big\{\text{Tr}^{(n)}\tau_p^{[B_{p-1}F_{p-1}]} f\Big\}\,(\tau_p^{[R_pB_p]})^{-1}\;\rho_{R_pB_p}^{(\text{mm})\;\otimes n} + \cdots\,, \tag{3.4}$$

for some f in the replicated Hilbert space. So each average over U_p comes with a sum over an element of the symmetric group $\tau_p \in S_n$. In the above, $\rho_{R_pB_p}^{(\text{mm})}$ is the maximally mixed state of $\mathcal{H}_{R_p}\otimes\mathcal{H}_{B_p}$ as in (2.20), while for the refined model, it is the subspace with energy in the window Θ_{p-1} as in (2.21). The ellipsis stand for subleading corrections, suppressed by inverse powers of $d_{B_{p-1}}d_{F_{p-1}}$, that we will not keep track of in our analysis. $\text{Tr}^{(n)}$ is the trace defined on the replicated Hilbert space.

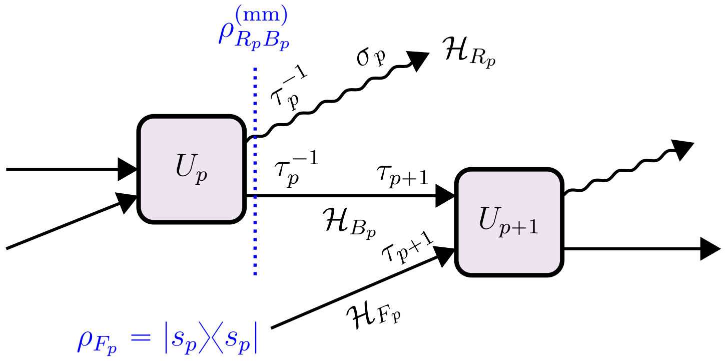

Figure 3. Assembling the ingredients for the building block in (3.6).

Applying (3.4) for all p, it becomes apparent that the average of (3.2) breaks up into a set of building blocks:

$$\overline{e^{(1-n)S^{(n)}(A)}} = \sum_{\tau_1,\dots,\tau_N \in S_n} \mathcal{Z}_1 \cdots \mathcal{Z}_N\,, \tag{3.5}$$

where

$$\mathcal{Z}_p = \mathrm{Tr}^{(n)} \sigma_p^{[R_p]} \tau_{p+1}^{[B_pF_p]} \big(\tau_p^{[R_pB_p]}\big)^{-1} \big(\rho^{(\mathrm{mm})}_{R_pB_p} \otimes \rho^{\mathrm{sc}}_{F_p}\big)^{\otimes n}\,, \tag{3.6}$$

where $\rho^{\mathrm{sc}}_{F_p} = |s_p\rangle\langle s_p|$ is the semi-classical state of the infalling system F_p. In the last step $p = N$ this piece is missing, there is no F_N. The traces over $\mathcal{H}_{F_p}$ are trivial because the states $\rho^{\mathrm{sc}}_{F_p}$ are pure and so $\mathrm{Tr}^{(n)}(\sigma \rho^{\mathrm{sc}\otimes n}_{F_p}) = 1$, for any $\sigma \in S_n$. This includes the initial state in $\mathcal{H}_{F_0}$ that collapsed to form the black hole. Hence, the building block (3.6) can be written more simply as

$$\mathcal{Z}_p = \mathrm{Tr}^{(n)} \sigma_p^{[R_p]} \tau_{p+1}^{[B_p]} \big(\tau_p^{[R_pB_p]}\big)^{-1} \rho^{(\mathrm{mm})\,\otimes n}_{R_pB_p}\,. \tag{3.7}$$

The expression for the building block $\mathcal{Z}_p$ can also be interpreted in terms of the equilibration ansatz of [46] as an alternative to the unitary averages. In this interpretation, the pure state of the black hole at time t_{p-1} equilibrates over the next time step meaning that for certain observables it is indistinguishable from an equilibrium state, in this case precisely the maximally mixed state $\rho^{(\mathrm{mm})}_{R_pB_p}$ (2.20), or (2.21) in the refined model.

In the basic model, it is then straightforward to evaluate the building block (3.7),

$$\mathcal{Z}_p = \exp\Big[\big(k(\tau_{p+1}\tau_p^{-1}) - n\big)S_{\mathrm{BH}}(M_p) + \big(k(\sigma_p\tau_p^{-1}) - n\big)S_{\mathrm{rad}}(R_p)\Big]\,, \tag{3.8}$$

where $k(\sigma)$ is the number cycles of the element σ and with $S_{\mathrm{BH}}(M_p) = \log d_{B_p}$ and $S_{\mathrm{rad}}(R_p) = \log d_{R_p}$. Then plugging into (3.5) gives the final result

$$\overline{e^{(1-n)S^{(n)}(A)}} = \sum_{\tau_1,\dots,\tau_N \in S_n} e^{(1-n)S^{(n)}_{\{\tau_p\}}(A)}\,, \tag{3.9}$$

where we have defined

$$S^{(n)}_{\{\tau_p\}}(A) = \frac{1}{n-1}\sum_{p=1}^{N}\Big\{d(\tau_{p+1},\tau_p)S_{\text{BH}}(M_p) + d(\sigma_p,\tau_p)S_{\text{rad}}(R_p)\Big\}\,, \tag{3.10}$$

where $d(\sigma,\pi) = n - k(\sigma\pi^{-1})$ is the Cayley distance between elements of S_n.[7]

3.1 Refined model

The refined model is rather more complicated because of the need to enforce energy conservation. The Rényi entropies now involve a sum over both the energies $\mathcal{E}_{j_p}$ and the elements of the symmetric group τ_p,

$$\overline{e^{(1-n)S^{(n)}(A)}} = \sum_{\{j_p\}}\sum_{\{\tau_p\}\in S_n}\prod_{p=1}^{n}\mathcal{Z}_p = \sum_{\{j_p\}}\sum_{\{\tau_p\}\in S_n} e^{(1-n)S^{(n)}_{\{\tau_p\}}(A)} \tag{3.11}$$

where the building block is

$$\mathcal{Z}_p = \frac{d_{R_p}(\mathcal{E}_{j_p})^{k(\sigma_p\tau_p^{-1})}d_B(M_{p-1}+E_p-\mathcal{E}_{j_p})^{k(\tau_{p+1}\tau_p^{-1})}}{\Big(\sum_{j_p} d_{R_p}(\mathcal{E}_{j_p})d_B(M_{p-1}+E_p-\mathcal{E}_{j_p})\Big)^{n}}\,. \tag{3.12}$$

where E_p is the energy of the infalling system F_p. Note that the mass of the black hole depends implicitly on the energy of the radiation emitted up to that point

$$M_p = M_0 + \sum_{q=1}^{p}(E_q - \mathcal{E}_{j_q})\,, \tag{3.13}$$

a point that must be born in mind when we perform the saddle point approximation.

The denominator in (3.12) can be evaluated by a saddle point approximation where the sum is replaced by an integral over a continuous variable $\mathcal{E}_p$. In particular, the radiation can be described thermodynamically in the way summarized in appendix A and the entropy

$$\log d_{R_p}(\mathcal{E}) = 2\sqrt{\mu_p\mathcal{E}_p} \qquad \text{where} \qquad \mu_p = \frac{\pi c\Delta t_p}{12}\,. \tag{3.14}$$

Since the saddle point value of the energy is much smaller than the black hole mass, the saddle point equation is

$$\sqrt{\frac{\mu_p}{\mathcal{E}_p}} = -\frac{dS_{\text{BH}}(M_{p-1}+E_p-\mathcal{E}_p)}{d\mathcal{E}_p} \approx \frac{1}{T_p} \qquad \Longrightarrow \qquad \mathcal{E}_p = \mu_p T_p^2\,, \tag{3.15}$$

where T_p defined in (2.6) is precisely the temperature of the Hawking radiation R_p. The average energy of the radiation emitted $\mathcal{E}_p$ and the infalling energy E_p are assumed to be much smaller than the black hole mass. Hence, we have

$$\sum_{j_p} d_{R_p}(\mathcal{E}_{j_p})d_B(M_{p-1}+E_p-\mathcal{E}_{j_p}) \approx d_B(M_{p-1})e^{S_{\text{rad}}(R_p)/2+E_p/T_p}\,, \tag{3.16}$$

[7] Alternatively, the Cayley distance $d(\sigma,\pi)$ may be defined as the minimal number of transpositions required to go between σ and π.

where the saddle point value of the entropy is

$$S_{\rm rad}(R_p) = 2\mu_p T_p = \frac{\pi c \Delta t_p T_p}{6} \,. \tag{3.17}$$

This and $\mathcal{E}_p$ above are the familiar expressions for the entropy and energy of a volume Δt_p of a relativistic gas in $1+1$ dimensions in a volume $\mathscr{V} = \Delta t_p$ (as reviewed in appendix A). The saddle point approximation is, of course, just the conventional way of deriving the Legendre transformation between the internal energy and free energy in thermodynamics and is justified precisely because the spread in the energy is small (2.13).

For later use, note that

$$d_B(M_p) = d_B(M_{p-1} + E_p - \mathcal{E}_p) \approx d_B(M_{p-1} + E_p) e^{-S_{\rm rad}(R_p)/2} \tag{3.18}$$

and so

$$S_{\rm BH}(M_{p-1} + E_p) - S_{\rm BH}(M_p) = \frac{S_{\rm rad}(R_p)}{2} \,, \tag{3.19}$$

which is the familiar relation for a model of black hole evaporation in the s-wave approximation and with no back scattering (i.e. grey body factor). Note that it implies that the evaporation is irreversible.

We now proceed to evaluate the sums of the energies in (3.11) by similar saddle point approximations. After we replace the sums by integrals over $\mathcal{E}_p$, the exponent of the integrand is

$$\begin{aligned}(1-n) S^{(n)}_{\{\tau_p\}}(A) = \sum_{p=1}^{N} \Bigg\{ & 2(n - d(\sigma_p, \tau_p)) \sqrt{\mu_p \mathcal{E}_p} - d(\tau_{p+1}, \tau_p) S_{\rm BH}(M_0) \\ & - \left(n - \sum_{q=p}^{N} d(\tau_{q+1}, \tau_q) \right) \frac{\mathcal{E}_p}{T_p} - \sum_{q=p}^{N} d(\tau_{q+1}, \tau_q) \frac{E_p}{T_p} - \frac{n}{2} S_{\rm rad}(R_p) \Bigg\} \,.\end{aligned} \tag{3.20}$$

It is now simple to compute the saddle point equations for the energies $\mathcal{E}_p$. In the regime of slow evaporation we can ignore the $\mathcal{E}_p$ dependence of the temperatures T_p. The saddle point values are found to be

$$\mathcal{E}_p = \mu_p T_p^2 \left(\frac{n - d(\sigma_p, \tau_p)}{n - \sum_{q=p}^N d(\tau_{q+1}, \tau_q)} \right)^2 \,, \tag{3.21}$$

where for consistency the saddles must have

$$n > \sum_{q=1}^{N} d(\tau_{q+1}, \tau_q) \,. \tag{3.22}$$

The contribution of this saddle to the Rényi entropy is

$$\begin{aligned} S^{(n)}_{\{\tau_p\}}(A) = \frac{1}{n-1} \sum_{p=1}^{N} \Bigg\{ & d(\tau_{p+1}, \tau_p) S_{\rm BH}(M_0) + \sum_{q=p}^{N} d(\tau_{q+1}, \tau_q) \frac{E_p}{T_p} \\ & + \frac{1}{2} \left(n - \frac{(n - d(\sigma_p, \tau_p))^2}{n - \sum_{q=p}^N d(\tau_{q+1}, \tau_q)} \right) S_{\rm rad}(R_p) \Bigg\} \,. \end{aligned} \tag{3.23}$$

We can re-write this by noting that (3.19) implies

$$S_{\mathrm{BH}}(M_p) = S_{\mathrm{BH}}(M_0) + \sum_{q=1}^{p} \left(\frac{E_q}{T_q} - \frac{S_{\mathrm{rad}}(R_q)}{2} \right) , \tag{3.24}$$

as

$$\begin{aligned} S^{(n)}_{\{\tau_p\}}(A) = \frac{1}{n-1} \sum_{p=1}^{N} \Bigg\{ & d(\tau_{p+1}, \tau_p) S_{\mathrm{BH}}(M_p) \\ & + \frac{2nd(\sigma_p, \tau_p) - d(\sigma_p, \tau_p)^2 - \left(\sum_{q=p}^{N} d(\tau_{q+1}, \tau_q) \right)^2}{2 \left(n - \sum_{q=p}^{N} d(\tau_{q+1}, \tau_q) \right)} S_{\mathrm{rad}}(R_p) \Bigg\} , \end{aligned} \tag{3.25}$$

which is the refined model generalization of (3.10).

3.2 Relation to the island formalism

We interpret (3.9) as being a sum over saddles of the (Lorentzian) gravitational path integral in the semi-classical limit, labelled by the elements $\{\tau_p\}$. In this limit, the entropies $S_{\mathrm{BH}}(M_p)$ and $S_{\mathrm{rad}}(R_p)$ are very large. If we avoid the crossover regimes when saddles are degenerate, it turns out that only a much smaller number of terms can actually dominate in the sum, namely, those for which each τ_p, $p = 1, \ldots, N$, is equal to e or η only, the identity and cyclic permutations, respectively. This is proved in appendix B for the basic model. The $\{e, \eta\}$ dominance means that the saddles that dominate respect the $\mathbb{Z}_n$ cyclic symmetry of the replicas mirroring the symmetry of the replica wormholes of [4, 5], or, equivalently, we can interpret the average over unitaries to be equivalent to the average over baby universe states (see [29, 47]).

For the refined model, the discussion is very similar. Indeed each element in the energy sum in (3.11) behaves like a basic model, and therefore we can again invoke the fact that τ_p is dominated by $\tau_p \in \{e, \eta\}$, which will be valid as long as we are not in the vicinity of a crossover of saddles.[8]

The expression for the von Neumann entropy of our chosen subset $A \subset R \cup B$ is obtained from (3.10) and (3.25) in the limit $S(A) = \lim_{n \to 1} S^{(n)}(A)$ and has the form of a minimization problem over the 2^N choices $\tau_p \in \{e, \eta\}$. Indeed notice that when $\sigma, \pi \in \{e, \eta\}$, we can write

$$d(\sigma, \pi) = (n-1)(1 - \delta_{\sigma\pi}) , \tag{3.26}$$

which facilitates the evaluation of the Cayley distances in the $n \to 1$ limit of (3.10) and (3.25). In both models, the von Neumann entropy is given by

$$S(A) = \min_{\{\tau_p\}} S_{\{\tau_p\}}(A) = \min_{\{\tau_p\}} \left\{ \sum_{p=1}^{N} (1 - \delta_{\tau_{p+1}\tau_p}) S_{\mathrm{BH}}(M_p) + (1 - \delta_{\sigma_p \tau_p}) S_{\mathrm{rad}}(R_p) \right\} . \tag{3.27}$$

[8] Notice that we don't risk of having a crossover at every time step since we assumed that each energy window is small (2.13).

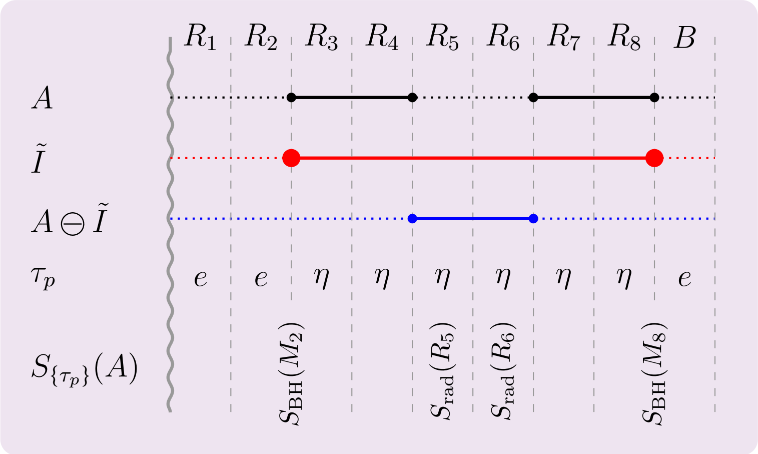

Figure 4. An example of a saddle for the model with $N = 8$ time steps, with some choice of the set $A = R_3 \cup R_4 \cup R_7 \cup R_8$, as shown, with an island-in-the-stream $\tilde{I}$. Note that $\partial\tilde{I} \subset \partial A$. The contributions to the entropy from each time step are shown and summing these up gives $S_I(A) = S_{\mathrm{BH}}(M_2) + S_{\mathrm{BH}}(M_8) + S_{\mathrm{rad}}(R_5 \cup R_6)$. Note that the last term is $S_{\mathrm{rad}}(A \ominus \tilde{I})$.

The resemblance of this equation to the QES formula described in the introduction for a slowly evaporating black hole (1.19) becomes more apparent if we set

$$S_I(A) \equiv S_{\{\tau_p\}}(A). \tag{3.28}$$

where I is defined in both models as

$$I = \bigcup_{p\in\Phi} \left(\overline{R}_p \cup F_{p-1}\right). \tag{3.29}$$

with $\Phi = \{p \mid \tau_p = \eta\}$. The I that minimizes (3.28) is called the 'entanglement island' or 'island', for short, will be denoted $I(A)$. Even if in principle we have 2^N possible saddles, most of them will not contribute since terms with $\tau_p \neq \tau_{p+1}$ are not favourable because of the black hole entropy being big. One can check that the only saddles that are not trivially suppressed are the one where τ_p changes in correspondence with a change in σ_p, which is an analog of the condition (1.17). See figure 4 for an example where $\Phi = \{3, 4, 5, 6, 7, 8\}$.

In order to make more transparent the identification of (3.27) with the QES formula (1.19) for the A that we have chosen, we can also notice that the second term is a discrete version of the continuum expression $S_{\mathrm{rad}}(A \ominus \tilde{I})$ where we identify the island-in-the-stream as the reflection of the island I in the horizon and then projected onto $\mathscr{I}^+$, so each $\overline{R}_p$ gets mapped to R_p:

$$\tilde{I} = \bigcup_{p\in\Phi} R_p. \tag{3.30}$$

On the other hand, the first term can be written in terms of the BH entropy at the outgoing EF coordinates of QES $u_{\partial I}$. We can then parametrize the entropy of the black hole with its mass at outgoing time u. Notice also that the infalling states in (3.29) are shifted by $p \to p - 1$. This is how the model accounts for the fact that infalling coordinate v of the QES are shifted relative to the outgoing coordinate u by the scrambling time (1.16), precisely the size of the time steps in the model.

In the next sections, we will enforce our definition of the entanglement island (3.29) studying when it is possible to reconstruct an unitary acting on the radiation, which is equivalent to the well known statement that the island is in the entanglement wedge of the radiation. Specifically, since the emitted radiation is in both the semi-classical and microscopic descriptions, we include it in its own entanglement wedge

$$\mathcal{W}(A) = I(A) \cup (A \cap R)\,. \tag{3.31}$$

Although we call $I(A)$ the island, strictly speaking, this only applies when subsets of $I(A)$ are separated from the rest of the entanglement wedge by QES.[9]

4 Information recovery and reconstruction

In this section, we consider the fate of an infalling system, Hayden and Preskill's diary for instance [36]. A version of this problem was considered for the basic (or BRU) model in section 7 of [28]. The purpose of this section is to extend this analysis to reconstruction on subsets of the radiation, which we verify is consistent with the QES formula discussed in section 3, and to extend this analysis to the refined model. It is worth noting that we employ a different method, as compared with [28], to show when reconstruction is possible by employing a replica trick to calculate the trace distance to find when certain decoupling conditions hold. Finally, we point out in section 4.1 how a different approach which uses standard bounds on the trace distance can not always be used in the refined model.

We will focus on a single system that falls in during the q^{th} time step. For simplicity, we will avoid the case that this is the last time step, in other words we will take $q < N$. The idea is to consider a family of infalling states $W|s_q\rangle$ for a unitary W and fixed state $|s_q\rangle \in F_q$. This gives a family of microscopic states $|\Psi(W)\rangle$. The physical question is, can the effect of the unitary W be achieved by a local action on the radiation or the black hole? This will inform us as to when the information in F_q has been teleported out of the black hole. More specifically, when can the action of W be *reconstructed* on $A = R$ or B, or a subset thereof, in the sense that there exists a unitary W_A acting on A such that

$$W_A|\Psi\rangle \overset{?}{=} |\Psi(W)\rangle\,. \tag{4.1}$$

This is the state-specific notion of reconstruction described in [28]. The above implies that W_A acts on the reduced state on A via the adjoint action

$$\rho_A(W) = W_A \rho_A W_A^\dagger\,, \tag{4.2}$$

while the reduced state on the complement $\overline{A}$ is invariant

$$\rho_{\overline{A}}(W) = \rho_{\overline{A}}\,. \tag{4.3}$$

[9] For example, when $A = B$, the black hole before the Page time has $I(B) = \mathcal{W}(B) = \overline{R} \cup F$ which is not an island in the strict sense.

In fact this *decoupling condition* on $\overline{A}$ implies the existence of W_A in (4.1). This can be seen using the Schmidt decomposition. The decoupling condition implies that if $|\Psi\rangle = \sum_j \sqrt{p_j}|j\rangle_A|j\rangle_{\overline{A}}$ then $|\Psi(W)\rangle = \sum_j \sqrt{p_j}|j\rangle'_A|j\rangle_{\overline{A}}$. It follows that $W_A = \sum_j |j\rangle'_A\langle j|$ acting on the subspace of $\mathcal{H}_A$ spanned by the Schmidt states $|j\rangle_A$ although it can be extended to a unitary acting on $\mathcal{H}_A$. Acting within the subspace, we can write explicitly,

$$W_A = \mathrm{Tr}_{\overline{A}}|\Psi(W)\rangle\langle\Psi|\rho_{\overline{A}}^{-1}\,. \tag{4.4}$$

The Schmidt basis states depend implicitly on the infalling state $|s_q\rangle$ and so the construction of W is 'state dependent' in this sense. It is an interesting question if the construction can be extended to any operator acting on any state of the infalling system in $\mathcal{H}_{F_q}$ and thereby be state independent, at least in this limited sense. In fact, the construction above can be seen as a special case of the Petz map and, indeed, there is a more general state-independent construction [48].

We cannot expect the conditions (4.1) and (4.3) to hold exactly and approximate forms of these conditions are formulated in [28]. However, we will work to leading order in the semi-classical limit and we will not need these approximate forms in our analysis.

The decoupling condition is therefore key to reconstructing that action of W on either the radiation or the black hole. Hence, we need to calculate the difference between the states $\rho_A(W)$ and ρ_A. This can be measured by the trace norm[10] difference $\|\sigma-\rho\|_1$ or the quantum fidelity $f(\sigma,\rho)$. Both are tractable in our models when averaged over the unitary evolution to leading order in the semi-classical limit where they can be computed using the replica method and an analytic continuation. For the trace norm difference, we take an even number of replicas and then take an analytic continuation,

$$\|\sigma-\rho\|_1 = \mathrm{Tr}\sqrt{(\sigma-\rho)^2} = \lim_{n\to\frac{1}{2}} \mathrm{Tr}^{(2n)}\eta\big(\sigma-\rho\big)^{\otimes 2n} \tag{4.5}$$

and similarly for the quantum fidelity,

$$f(\sigma,\rho) \equiv \mathrm{Tr}\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}} = \lim_{n\to\frac{1}{2}} \mathrm{Tr}^{(2n)}\eta\big(\sigma\otimes\rho\big)^{\otimes n}\,. \tag{4.6}$$

In our context, there is a subtlety in that the analytic continuations must be taken *after* the semi-classical limit has picked out a dominant saddle otherwise saddles would become degenerate. We should also emphasize that what we are actually calculating are the unitary averages of the replica expressions before taking the limits $n\to\frac{1}{2}$. This is in the same spirit as calculating the averages the exponents of the Rényi entropies as in (3.5) before taking the limit $n\to 1$ to recover the von Neumann entropy. In section 4.1 we compute an upper bound on the trace norm which does not require the $n\to\frac{1}{2}$ limit.

Let us compute the average of the trace difference in (4.5). The computation is similiar to that of the Rényi entropy via $\mathrm{Tr}\rho_A^n$. In fact, since W acts locally on $\mathcal{H}_{F_q}$, only the q^{th} time step is modified:

$$\begin{aligned}\mathcal{Z}_q &\longrightarrow \mathrm{Tr}^{(2n)}\sigma_q^{[R_q]}\tau_{q+1}^{[B_qF_q]}\big(\tau_q^{[R_qB_q]}\big)^{-1}\big(\rho^{(\mathrm{mm})}_{R_qB_q}\otimes\big(\rho^{\mathrm{sc}}_{F_q}(W)-\rho^{\mathrm{sc}}_{F_q}\big)\big)^{\otimes 2n}\\ &= \mathcal{Z}_q\,\mathrm{Tr}^{(2n)}\tau_{q+1}\big(\rho^{\mathrm{sc}}_{F_q}(W)-\rho^{\mathrm{sc}}_{F_q}\big)^{\otimes 2n}\,,\end{aligned} \tag{4.7}$$

[10] For Hermitian operators the trace norm is equal to $\|\mathcal{O}\|_1 = \sum_j|\lambda_j|$, where λ_j are the eigenvalues of $\mathcal{O}$.

where we separated out the trace over the replicas of F_q where W acts and the quantity $\mathcal{Z}_q$ is the original quantity in the entropy calculation defined in (3.7). The contribution from the other time steps $p \neq q$ are precisely as for the entropy (3.7). Hence, assembling all the pieces gives

$$\overline{\left\|\rho_A(W) - \rho_A\right\|_1} = \lim_{n\to\frac{1}{2}} \sum_{\tau_1,\ldots,\tau_N \subset \{e,\eta\}} e^{(1-2n)S^{(2n)}_{\{\tau_p\}}(A)}\, \mathrm{Tr}^{(2n)} \tau_{q+1}\big(\rho^{\mathrm{sc}}_{F_q}(W) - \rho^{\mathrm{sc}}_{F_q}\big)^{\otimes 2n} . \tag{4.8}$$

Now we have to be careful to take the semi-classical limit before taking the analytic continuation $n \to \frac{1}{2}$. The semi-classical limit picks out a dominant term in the sum over the elements τ_p and, in particular, fixes τ_{q+1}. Hence,

$$\overline{\left\|\rho_A(W) - \rho_A\right\|_1} = \lim_{n\to\frac{1}{2}} \mathrm{Tr}^{(2n)} \tau_{q+1}\big(\rho^{\mathrm{sc}}_{F_q}(W) - \rho^{\mathrm{sc}}_{F_q}\big)^{\otimes 2n} \tag{4.9}$$

One can follow the same steps for the average of the quantum fidelity. Once again the contribution comes entirely from the q^{th} time step which is modified as

$$\begin{aligned} \mathcal{Z}_q &\longrightarrow \mathrm{Tr}^{(2n)} \sigma_q^{[R_q]} \tau_{q+1}^{[B_qF_q]} \big(\tau_q^{[R_qB_q]}\big)^{-1} \big(\rho^{(\mathrm{mm})}_{R_qB_q}\big)^{\otimes 2n} \otimes \big(\rho^{\mathrm{sc}}_{F_q}(W) \otimes \rho^{\mathrm{sc}}_{F_q}\big)^{\otimes n} \\ &= \mathcal{Z}_q\, \mathrm{Tr}^{(2n)} \tau_{q+1}\big(\rho^{\mathrm{sc}}_{F_q}(W) \otimes \rho^{\mathrm{sc}}_{F_q}\big)^{\otimes n} , \end{aligned} \tag{4.10}$$

leading to

$$\overline{f(\rho_A(W), \rho_A)} = \lim_{n\to\frac{1}{2}} \mathrm{Tr}^{(2n)} \tau_{q+1}\big(\rho^{\mathrm{sc}}_{F_q}(W) \otimes \rho^{\mathrm{sc}}_{F_q}\big)^{\otimes n} \tag{4.11}$$

Let us now evaluate our results above. When $F_q \notin \mathcal{W}(A)$, it follows that the dominant saddle has $\tau_{q+1} = e$. For the trace norm difference (4.9), this gives an expression that is clearly seen to vanish

$$\begin{aligned} \overline{\left\|\rho_A(W) - \rho_A\right\|_1} &= \lim_{n\to\frac{1}{2}} \mathrm{Tr}^{(2n)} \big(\rho^{\mathrm{sc}}_{F_q}(W) - \rho^{\mathrm{sc}}_{F_q}\big)^{\otimes 2n} \\ &= \big|\mathrm{Tr}\big(\rho^{\mathrm{sc}}_{F_q}(W) - \rho^{\mathrm{sc}}_{F_q}\big)\big| = 0 . \end{aligned} \tag{4.12}$$

This proves the decoupling condition in terms of the trace norm. On the other hand, for the fidelity (4.11),[11]

$$\begin{aligned} \overline{f(\rho_A(W), \rho_A)} &= \lim_{n\to\frac{1}{2}} \mathrm{Tr}^{(2n)} \big(\rho^{\mathrm{sc}}_{F_q}(W) \otimes \rho^{\mathrm{sc}}_{F_q}\big)^{\otimes n} \\ &= \sqrt{\mathrm{Tr}\rho^{\mathrm{sc}}_{F_q}(W)\, \mathrm{Tr}\rho^{\mathrm{sc}}_{F_q}} = 1 , \end{aligned} \tag{4.13}$$

which is another expression of decoupling. Note that, if the trace norm difference of two states vanishes, then they must have unit quantum fidelity and $\rho_A(W) = \rho_A$.

[11] The fidelity plays an important role in quantum hypothesis testing, which is the task of making a measurement to distinguish between two quantum states given that the actual state is one of them. The fidelity bounds the error on the optimal measurement. We expect that the corrections to (4.13) are non-perturbatively suppressed in the semi-classical limit, as in [57, 58]. If so, this would imply that whilst it is not possible to distinguish the two states given a single copy of the state, it will be possible given sufficiently many copies of the state.

On the other hand, when $F_q \in \mathcal{W}(A)$, the element $\tau_{q+1} = \eta$ and the trace norm difference (4.9) is

$$\begin{aligned}\overline{\big\|\rho_A(W) - \rho_A\big\|_1} &= \lim_{n\to\frac{1}{2}} \operatorname{Tr}^{(2n)} \eta \big(\rho^{\text{sc}}_{F_q}(W) - \rho^{\text{sc}}_{F_q}\big)^{\otimes 2n} \\ &= \big\|\rho^{\text{sc}}_{F_q}(W) - \rho^{\text{sc}}_{F_q}\big\|_1 .\end{aligned} \tag{4.14}$$

For the fidelity, we have a similar relation to the semi-classical state

$$\begin{aligned}\overline{f(\rho_A(W), \rho_A)} &= \lim_{n\to\frac{1}{2}} \operatorname{Tr}^{(2n)} \eta \big(\rho^{\text{sc}}_{F_q}(W) \otimes \rho^{\text{sc}}_{F_q}\big)^{\otimes n} \\ &= f\big(\rho^{\text{sc}}_{F_q}(W), \rho^{\text{sc}}_{F_q}\big) .\end{aligned} \tag{4.15}$$

Let us take stock of the results and, in particular, relate them to the *state reconstruction formula* of [49]. This states that if there are two microscopic states ρ_A and σ_A such that the semi-classical saddles that dominate $\operatorname{Tr}(\rho_A^{2n})$ and $\operatorname{Tr}(\sigma_A^{2n})$ are the same (and preserve the $\mathbb{Z}_n$ symmetry of the replicas) then[12]

$$\big\|\rho_A - \sigma_A\big\|_1 = \big\|\rho^{\text{sc}}_{\mathcal{W}(A)} - \sigma^{\text{sc}}_{\mathcal{W}(A)}\big\|_1 , \tag{4.16}$$

up to $\mathscr{O}(G)$ corrections, where $\mathcal{W}(A)$ is the entanglement wedge of A. The fact that the map V preserves the trace norm difference is on the same footing as the preservation of the relative entropy [5, 25, 50].

To relate this to our analysis, we identify $\sigma_R = \rho_R(W)$. The saddles associated to $\operatorname{Tr}(\rho_A^{2n})$ and $\operatorname{Tr}(\sigma_A^{2n})$ are the ones that determine the Rényi entropies and are therefore associated to the set of elements τ_p, $p = 1, \ldots, N$. The fact that they both have the same saddle is ensured by the fact that W only acts on a small subset of the infalling modes and so cannot alter the dominant saddle.

Let us consider our results for the case $A = R$. Before the Page time, $\mathcal{W}(R) = R$ and so $F_q \notin \mathcal{W}(R)$ and the formula (4.16) implies

$$\big\|\rho_R(W) - \rho_R\big\|_1 = 0 , \tag{4.17}$$

which is the decoupling condition (4.12) with $A = R$. This means that W can be reconstructed on B. On the other hand, after the Page time, the entanglement wedge $\mathcal{W}(R) = R \cup I(R)$, so $F_q \in \mathcal{W}(R)$, since the island $I(R)$ contains the outgoing and infalling modes $I(R) = \overline{R} \cup F$ since it lies very close behind the horizon. Hence, (4.16) implies

$$\big\|\rho_R(W) - \rho_R\big\|_1 = \big\|\rho^{\text{sc}}_{R\overline{R}F}(W) - \rho^{\text{sc}}_{R\overline{R}F}\big\|_1 = \big\|\rho^{\text{sc}}_{F_q}(W) - \rho^{\text{sc}}_{F_q}\big\|_1 , \tag{4.18}$$

which is (4.14) with $A = R$. We will see shortly that this is the case when W can be reconstructed on R because B decouples.

[12] We have stated the formula in a slightly more general way to include the case when A is any subset of the radiation plus the black hole rather than all the radiation as considered in [49]. The condition for $\mathbb{Z}_n$ symmetry is satisfied by our saddles which involve only the elements e or η of S_n.

Now consider the case $A = B$. After the Page time, $\mathcal{W}(B) = \varnothing$ and so (4.16) predicts decoupling as we found in (4.12). This occurs at the same time as (4.18) which makes perfect sense as W can be reconstructed on R. On the other hand, before the Page time, $\mathcal{W}(B) = \overline{R} \cup F$, and so (4.16) gives

$$\left\|\rho_B(W) - \rho_B\right\|_1 = \left\|\rho^{\text{sc}}_{RF}(W) - \rho^{\text{sc}}_{RF}\right\|_1 = \left\|\rho^{\text{sc}}_{F_q}(W) - \rho^{\text{sc}}_{F_q}\right\|_1 . \tag{4.19}$$

But this is precisely (4.14) for $A = B$. This is also when R decouples and so W can be reconstructed on B. So once again we find precise agreement between our averaged results and the formula (4.16).

4.1 Bounding the trace norm

The condition for decoupling is that the averaged trace norm difference between $\rho_A(W)$ and ρ_A vanishes in the leading order saddle (4.12). But this is derived with the limits in a particular order, first the semi-classical limit picking out a particular saddle and then in the replica limit $n \to \frac{1}{2}$. Can we trust this? In fact there is standard way to bound the averaged trace norm difference,

$$\overline{\left\|\rho_A(W) - \rho_A\right\|_1} \leqslant \sqrt{d_A\, \overline{\text{Tr}(\rho_A(W) - \rho_A)^2}} . \tag{4.20}$$

We can evaluate the right-hand side, at least in the case that the subsystem A is finite dimensional. Note that this seems to exclude $A = R$, the radiation, in the refined model as it has infinite dimension. The average on the right-hand side is just the right-hand side of (4.8) with $n \to 1$, so

$$\overline{\text{Tr}(\rho_A(W) - \rho_A)^2} = \sum_{\tau_1,\ldots,\tau_N \subset \{e,\eta\}} e^{-S^{(2)}_{\{\tau_p\}}(A)}\, \text{Tr}^{(2)} \tau_{q+1} \big(\rho^{\text{sc}}_{F_q}(W) - \rho^{\text{sc}}_{F_q}\big)^{\otimes 2} . \tag{4.21}$$

If we consider $A = B$, so $d_A \sim e^{S_{\text{BH}}(M)}$, and after the Page time, the sum in (4.21) is dominated by the term with $\tau_p = \eta$ for which $S^{(2)}_{\{\eta\}}(B) = \alpha S_{\text{rad}}(R)$, where $\alpha = 1$ for the basic model and $\alpha = \frac{3}{4}$, for the refined model.[13] Therefore we can bound the trace norm difference

$$\overline{\left\|\rho_A(W) - \rho_A\right\|_1} \lessapprox \mathscr{O}\left(e^{\frac{1}{2}S_{\text{BH}}(M) - \frac{\alpha}{2}S_{\text{rad}}(R)}\right) \ll 1 , \tag{4.22}$$

after the Page time when $S_{\text{rad}}(R) \gg S_{\text{BH}}(B)$.

5 Reconstruction of the Hawking partners

In this section, we consider reconstruction for the Hawking partners which semi-classically are behind the horizon and part of the black hole. This problem was considered in the static model of [28], which essentially represents the holographic map at a fixed time. Therefore, it differs from the dynamical models considered in this paper. In addition, we

[13] The latter follows from (3.25) with $\tau_p = \eta$, $p = 1, \ldots, N+1$ and $\sigma_p = e$ giving $d(\tau_{p+1}, \tau_p) = 0$ and $d(\sigma_p, \tau_p) = n - 1$ giving $S^{(n)}_{\{\eta\}}(B) = (n+1)S_{\text{rad}}(R)/(2n)$. This is the Rényi entropy of the radiation (see appendix A) and then taking $n = 2$ gives $\frac{3}{4}S_{\text{rad}}(R)$.

note that section 8.2 of [28] addressed a different yet related problem concerning the ability to distinguish certain interior states.

Conceptually the discussion is very similar to the reconstruction of the infalling system in the last section but the technical details are rather different. The idea is to consider a unitary operator on the Hawking partners $\overline{R}$ and ask if it is possible to reconstruct this on some $A \subset R \cup B$, i.e.

$$|\Psi(W)\rangle \overset{?}{=} W_A|\Psi\rangle\,. \tag{5.1}$$

As in section 4 the condition for such a reconstruction is the decoupling condition for the complement

$$\rho_{\overline{A}}(W) = \rho_{\overline{A}}\,, \tag{5.2}$$

which can be analysed by calculating the trace norm difference or quantum fidelity.

In order to proceed, it is useful to deploy the following trick. Exploiting the entanglement between $\overline{R}$ and R, we can write the action of W on the semi-classical state as the action of an operator $\widetilde{W}$ on R:

$$W|\psi\rangle = \widetilde{W}|\psi\rangle\,, \tag{5.3}$$

where

$$\widetilde{W} = (\rho_R^{\text{sc}})^{1/2} W^T (\rho_R^{\text{sc}})^{-1/2}\,. \tag{5.4}$$

We remark that $\widetilde{W}$ is not unitary so it is not a physically realizable local action on the radiation. It then follows that the reduced state on R is invariant under adjoint action by $\widetilde{W}$,

$$\rho_R^{\text{sc}} \longrightarrow \widetilde{W}\rho_R^{\text{sc}}\widetilde{W}^\dagger = (\rho_R^{\text{sc}})^{1/2}\big(W^\dagger W\big)^*(\rho_R^{\text{sc}})^{1/2} = \rho_R^{\text{sc}}\,, \tag{5.5}$$

as it must be by locality: the action of W on $\overline{R}$ cannot change the state of R.

We now compute the trace difference and quantum fidelity of the two states $\rho_A(W)$ and ρ_A using the replica method following the same steps as in section 4. For simplicity, we will take W to act on just one of the subsets of partner modes $\overline{R}_q$. We can then use (5.3) to write the action on the Hawking modes R_q by switching $W \to \widetilde{W}$. As for the infalling system, the only effect of W is on the q^{th} time step. For the trace norm difference, this time step is modified as

$$\mathcal{Z}_q \longrightarrow \text{Tr}^{(2n)}\sigma_q^{[R_q]}\tau_{q+1}^{[B_q]}\big(\text{Ad}_{\widetilde{W}}-1\big)^{\otimes 2n}\big(\tau_q^{[R_qB_q]}\big)^{-1}\rho_{R_qB_q}^{(\text{mm})\ \otimes 2n}\,, \tag{5.6}$$

where $\text{Ad}_{\widetilde{W}}$ is the adjoint action of $\widetilde{W}$ on $\rho_{R_qB_q}^{(\text{mm})}$. We now assume that the saddle that dominates the entropy, and therefore the trace norm difference, has $\tau_{q+1} = \tau_q$. This means that $\overline{R}_q$ is not just before a QES. One can view this as avoiding an edge effect created by having a discrete model. In that case, we can perform the trace over B_q to give the semi-classical state $\rho_{R_q}^{\text{sc}} = \text{Tr}_{B_q}\rho_{R_qB_q}^{(\text{mm})}$:

$$\mathcal{Z}_q \longrightarrow \text{Tr}^{(2n)}\sigma_q\big(\text{Ad}_{\widetilde{W}}-1\big)^{\otimes 2n}\tau_q^{-1}\,\rho_{R_q}^{\text{sc}\ \otimes 2n}\,. \tag{5.7}$$

where in the second line we used the fact that all relevant saddles have $\tau_{q+1} = \tau_q$ and $\rho_{R_q}^{\text{sc}} = \text{Tr}_{B_q}\rho_{R_qB_q}^{(\text{mm})}$. Following the same steps as in section 4, and in particular taking the semi-classical limit before the analytic continuation in n, gives

$$\overline{\big\|\rho_A(W)-\rho_A\big\|_1} = \lim_{n\to\frac{1}{2}} \mathrm{Tr}^{(2n)}\sigma_q\big(\mathrm{Ad}_{\widetilde{W}}-1\big)^{\otimes 2n}\tau_q^{-1}\,\rho_{R_q}^{\mathrm{sc}\;\otimes 2n} \tag{5.8}$$

where τ_q is determined by the saddle that dominates the entropy. Similarly, for the quantum fidelity

$$\overline{f(\rho_A(W),\rho_A)} = \lim_{n\to\frac{1}{2}} \mathrm{Tr}^{(2n)}\sigma_q\big(\mathrm{Ad}_{\widetilde{W}}\otimes 1\big)^{\otimes n}\tau_q^{-1}\,\rho_{R_q}^{\mathrm{sc}\;\otimes 2n} \tag{5.9}$$

When $\overline{R}_q \notin \mathcal{W}(A)$ the dominant saddle has $\tau_q = e$ and then the trace norm difference is

$$\overline{\big\|\rho_A(W)-\rho_A\big\|_1} = \lim_{n\to\frac{1}{2}} \mathrm{Tr}^{(2n)}\sigma_q\big(\widetilde{W}\rho^{\mathrm{sc}}_{R_q}\widetilde{W}^\dagger-\rho^{\mathrm{sc}}_{R_q}\big)^{2n} = 0\,, \tag{5.10}$$

using the invariance (5.5). We can repeat the analysis for the fidelity,

$$\begin{aligned}\overline{f(\rho_A(W),\rho_A)} &= \lim_{n\to\frac{1}{2}} \mathrm{Tr}^{(2n)}\sigma_q\Big(\widetilde{W}\rho^{\mathrm{sc}}_{R_q}\widetilde{W}^\dagger\otimes\rho^{\mathrm{sc}}_{R_q}\Big)^{\otimes n}\\ &= \lim_{n\to\frac{1}{2}} \mathrm{Tr}^{(2n)}\sigma_q\,\rho_{R_q}^{\mathrm{sc}\;\otimes 2n} = 1\,.\end{aligned} \tag{5.11}$$

So decoupling occurs when the partners $\overline{R}_q$ do not lie in the entanglement wedge of A. Under these circumstances, W can be reconstructed on the complement $\overline{A}$.

On the other hand, when $\overline{R}_q \in \mathcal{W}(A)$, the appropriate saddle has $\tau_q = \eta$ and (5.8) becomes

$$\overline{\big\|\rho_A(W)-\rho_A\big\|_1} = \lim_{n\to\frac{1}{2}} \mathrm{Tr}^{(2n)}\sigma_q\big(\mathrm{Ad}_{\widetilde{W}}-1\big)^{\otimes 2n}\eta^{-1}\rho_{R_q}^{\mathrm{sc}\;\otimes 2n}\,. \tag{5.12}$$

We can now consider this for particular choices for A. For the case $A = R$, so after the Page time, then $\sigma_q = \eta$, and the above becomes

$$\begin{aligned}\overline{\big\|\rho_R(W)-\rho_R\big\|_1} &= 2\sqrt{1-\big|\mathrm{Tr}\big(\rho^{\mathrm{sc}}_{R_q}W^T\big)\big|^2}\\ &= \big\|\rho^{\mathrm{sc}}_{R\overline{R}}(W)-\rho^{\mathrm{sc}}_{R\overline{R}}\big\|_1\,.\end{aligned} \tag{5.13}$$

For the case $A = B$, so before the Page time, $\sigma_q = e$, we have

$$\begin{aligned}\overline{\big\|\rho_B(W)-\rho_B\big\|_1} &= \lim_{n\to\frac{1}{2}} \mathrm{Tr}\big(W^*\rho^{\mathrm{sc}}_{R_q}W^T-\rho^{\mathrm{sc}}_{R_q}\big)^{2n}\\ &= \lim_{n\to\frac{1}{2}} \mathrm{Tr}\big(W\rho^{\mathrm{sc}}_{\overline{R}_q}W^\dagger-\rho^{\mathrm{sc}}_{R_q}\big)^{2n}\\ &= \big\|\rho^{\mathrm{sc}}_{\overline{R}}(W)-\rho^{\mathrm{sc}}_{\overline{R}}\big\|_1\,.\end{aligned} \tag{5.14}$$

Note that (5.14) is not the same as (5.13) because R is entangled with $\overline{R}$.

We can also consider the quantum fidelity. For $A = R$ (after the Page time),

$$\begin{aligned}\overline{f(\rho_R(W),\rho_R)} &= \lim_{n\to\frac{1}{2}} \mathrm{Tr}^{(2n)}\eta\big(\mathrm{Ad}_{\widetilde{W}}\otimes 1\big)^{\otimes 2n}\eta^{-1}\rho_{R_q}^{\mathrm{sc}\;\otimes 2n}\\ &= \big|\mathrm{Tr}(\rho^{\mathrm{sc}}_{R_q}W^T)\big| = f(\rho^{\mathrm{sc}}_{R\overline{R}}(W),\rho^{\mathrm{sc}}_{R\overline{R}})\end{aligned} \tag{5.15}$$

and for $A = B$ (before the Page time),

$$\begin{aligned}\overline{f(\rho_B(W),\rho_B)} &= \lim_{n\to\frac{1}{2}} \mathrm{Tr}^{(2n)}\big(\mathrm{Ad}_{\widetilde{W}}\otimes 1\big)^{\otimes 2n}\eta^{-1}\rho_{R_q}^{\mathrm{sc}\ \otimes 2n}\\ &= \lim_{n\to\frac{1}{2}} \mathrm{Tr}^{(2n)}\eta\big(W^*\rho^{\mathrm{sc}}_{R_q}W^T\otimes\rho^{\mathrm{sc}}_{R_q}\big)^n \\ &= f(\rho^{\mathrm{sc}}_{\overline{R}}(W),\rho^{\mathrm{sc}}_{\overline{R}})\,.\end{aligned}\tag{5.16}$$

These expressions are close cousins of the expressions for the trace norm difference in (5.13) and (5.14).

Once again, let us compare our results to the formula (4.16) of [49]. Firstly, let us compare the microscopic states $\rho_R(W)$ and ρ_R. Before the Page time, $\overline{R}_q \notin \mathcal{W}(R)$ and so (4.16) implies $\|\rho_R(W)-\rho_R\|_1 = 0$. After the Page time, $\overline{R}_q \in \mathcal{W}(R)$ and so (4.16) implies

$$\|\rho_R(W)-\rho_R\|_1 = \|\rho^{\mathrm{sc}}_{R\overline{R}F}(W)-\rho^{\mathrm{sc}}_{R\overline{R}F}\|_1\,,\tag{5.17}$$

which is precisely (5.13) because F is not entangled with $R\cup\overline{R}$.

Now we turn to the states $\rho_B(W)$ and ρ_B. In this case, after the Page time, $\mathcal{W}(B) = \varnothing$ and so (4.16) implies $\|\rho_B(W)-\rho_B\|_1 = 0$. On the other hand, before the Page time, $\mathcal{W}(B) = \overline{R}\cup F$, and so (4.16) implies

$$\|\rho_R(W)-\rho_R\|_1 = \|\rho^{\mathrm{sc}}_{\overline{R}F}(W)-\rho^{\mathrm{sc}}_{\overline{R}F}\|_1\,.\tag{5.18}$$

This is precisely (5.14) because F is not entangled with $\overline{R}$.

6 Discussion

We have defined a simple model which accounts for energy conservation that captures the information flow of an evaporating black hole and have seen that many of the features of the model proposed in [28] can be extended to this refined model. Unitarity is built in and this manifests at the level of the entropy of the radiation in the form of a discrete version of the QES variational problem. The model then allowed us to investigate in detail entanglement wedge reconstruction for a system that falls into the black hole and also for local actions on the Hawking partners. Our main contributions here were to study these problems in a model which generalised the model of [28] and also to consider when reconstruction is possible not just on the radiation or the black hole, but a subset thereof. The model reproduces the properties of the holographic map that have been proposed in [28]; namely, the map acts trivially on the outgoing radiation and non-isometrically on the black hole. This latter fact manifests the fact that the Hilbert space of an old black hole is not large enough to host all the Hawking partners of the semi-classical state. Something must give, the map is non-isometric and as a result the Hawking partners have been teleported out into the radiation as subtle features of the microscopic state of R. In a sense, when a black hole is past the Page time according to an external observer, its inside has been squeezed out into the radiation leaving only a small region between the horizon and the QES that could be thought of as being part of the black hole.

Although the proposal of [28] has clarified certain issues, much remains to be understood. Of principal interest is the fate of an infalling system. According to these models, an infalling system begins to be scrambled immediately. In fact, the infalling system will soon enter the entanglement wedge of a late-time observer who collects all the radiation, since the QES is very close up behind the horizon meaning that the information of the infalling observer is in the radiation available to the late-time observer. Is this compatible with the idea that the infalling system experiences a smooth internal geometry after horizon crossing? We have argued at the microscopic level, the state of the radiation is not the inertial vacuum in the neighbourhood of the horizon but perhaps the infalling system sees effectively a smooth geometry and being thermalized takes some time. This would support an idea previously presented in [51] and analysed in the basic model in [28]. The situation seems quite analogous to the same questions for the fuzzball paradigm in string theory [52, 53]. In that context, it is argued that a macroscopic (i.e. high energy) infalling system would take time to be thermalized as it falls into the fuzzball. In a proposal known as *fuzzball complementarity*, the high energy infallling system would not resolve the subtle structure of the microscopic state and effectively average it to see a smooth geometry. It seems plausible that the same mechanism is at work here, if an observer cannot resolve the fine details of ρ_R maybe it effectively experiences the average $\overline{\rho_R} = \rho_R^{\text{sc}}$, precisely the semi-classical state and a smooth horizon, at least for a while.

Acknowledgments

TJH, AL and SPK acknowledge support from STFC grant ST/T000813/1. NT and ZG acknowledge the support of an STFC Studentship. AL has also received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 804305).

A Thermodynamics of free fields

Consider a set of free fields in $1+1$ dimensions. We will consider just the right-moving modes. The canonical partition function of a single mode of energy ω is equal to

$$\mathcal{Z} = \sum_{p=0}^{\infty} e^{-p\omega/T} = \frac{1}{1-e^{-\omega/T}}\,, \qquad \sum_{p=0}^{1} e^{-p\omega/T} = 1 + e^{-\omega T}\,, \tag{A.1}$$

for a scalar and spinor field, respectively. Summing over modes in a volume $\mathscr{V}$ and assuming there are $N = c, 2c$ fields for bosons/fermions, gives the free energy

$$f = \pm N\mathscr{V}T \int_0^\infty \frac{d\omega}{2\pi} \log(1 \mp e^{-\omega/T}) = -\frac{\pi c\mathscr{V}T^2}{12}\,. \tag{A.2}$$

The average energy

$$\mathcal{E} = N\mathscr{V} \int_0^\infty \frac{d\omega}{2\pi} \frac{\omega}{e^{\omega/T} \mp 1} = \frac{\pi c\mathscr{V}T^2}{12} \tag{A.3}$$

and the entropy

$$S_{\text{rad}} = N\mathscr{V} \int_0^\infty \frac{d\omega}{2\pi} \left\{ \frac{\omega}{T(e^{\omega/T} \mp 1)} \mp \log(1 \mp e^{-\omega/T}) \right\} = \frac{\pi c \mathscr{V} T}{6} \,. \tag{A.4}$$

We can also evaluate the Rényi entropes,

$$\begin{aligned} (1-n)S_{\text{rad}}^{(n)} &= N\mathscr{V} \int_0^\infty \frac{d\omega}{2\pi} \log \sum_{p=0}^{\infty,1} \left(\frac{e^{-p\omega/T}}{\mathcal{Z}} \right)^n \\ &= N\mathscr{V} \int_0^\infty \frac{d\omega}{2\pi} \big(\log \mathcal{Z}(T/n) - n \log \mathcal{Z}(T) \big) \,. \end{aligned} \tag{A.5}$$

Hence,

$$S_{\text{rad}}^{(n)} = \frac{nf(T) - nf(T/n)}{(1-n)T} = \frac{1+n}{n} \mu T = \frac{1+n}{2n} S_{\text{rad}} \,. \tag{A.6}$$

We will need to understand whether the relativistic gas can be described thermodynamically. We can solve for the entropy in terms of the entropy, $S_{\text{rad}} = 2\sqrt{\mu \mathcal{E}}$, where $\mu = \pi c \mathscr{V}/12$. In the thermodynamic it should be possible to approximate the canonical partition function as a integral over a continuum set of states with energy $\mathcal{E}$ and density of states $e^{S_{\text{rad}}(\mathcal{E})}$, that is

$$\mathcal{Z} = e^{-f/T} = \int d\mathcal{E}\, e^{S_{\text{rad}}(\mathcal{E}) - \mathcal{E}/T} \,. \tag{A.7}$$

The thermodynamic limit can be understood as when the saddle point approximation of this integral is valid. The saddle point equation corresponds to the Legendre transformation between the internal energy and free energy:

$$f = \underset{\mathcal{E}}{\text{ext}} \big(\mathcal{E} - T S_{\text{rad}}(\mathcal{E}) \big) \,, \tag{A.8}$$

and has solution

$$\mathcal{E} = \mu T^2 \,, \tag{A.9}$$

for which the free energy

$$f = -\mu T^2 \,. \tag{A.10}$$

One can verify that these expressions are entirely consistent with (A.2) and (A.3). The saddle point approximation is valid in the limit that the spread in the energy around the saddle point $\Delta\mathcal{E} \ll \mathcal{E}$ which is the condition

$$\frac{\Delta \mathcal{E}}{\mathcal{E}} \sim \frac{1}{\sqrt{S_{\text{rad}}}} \ll 1 \,. \tag{A.11}$$

So when $S_{\text{rad}} \gg 1$, the gas can be described thermoydnamically.

B Dominant saddles

In the model, we encounter sums over elements of the symmetric group of the form (3.5). This motivates analysing a sum of the form

$$\mathcal{Z}(n) = \sum_{\sigma \in S_n} d_1^{-d(\sigma,\tau_1)} d_2^{-d(\sigma,\tau_2)} d_3^{-d(\sigma,\tau_3)} \,, \tag{B.1}$$

where $d_i \geqslant 1$, $\tau_i \in S_n$ and $d(\sigma, \pi)$ is the Cayley distance between elements of S_n. This is equal to

$$d(\sigma, \pi) = n - k(\sigma\pi^{-1}) \,, \tag{B.2}$$

where $k(\sigma)$ is the number of cycles the make up σ, e.g. $k(e) = n$ and $k(\eta) = 1$.

We are interested in minimising the following 'free energy'

$$f(\sigma) = x_1 d(\sigma, \tau_1) + x_2 d(\sigma, \tau_2) + x_3 d(\sigma, \tau_3) \,, \tag{B.3}$$

where $x_i = \log d_i$. We first consider the permutations which minimise the free energy at the following special regions in the *phase diagram* (see figure 5), which we may parameterise by x_1/x_3 and x_2/x_3:

- for $x_1/x_3 \to 0$ and $x_2/x_3 \to 0$: $f(\sigma) \to x_3 d(\sigma, \tau_3)$ is minimised for $\sigma = \tau_3$.
- for $x_1/x_3 + x_2/x_3 = 1$: $f(\sigma) = x_1 \left(d(\tau_1, \sigma) + d(\sigma, \tau_3)\right) + x_2 \left(d(\tau_2, \sigma) + d(\sigma, \tau_3)\right)$ is minimised for $\sigma \in \Gamma(\tau_1, \tau_3) \cap \Gamma(\tau_2, \tau_3)$. Here, $\Gamma(\tau_i, \tau_j)$ denotes the set of permutations σ which saturate the triangle inequality $d(\tau_i, \sigma) + d(\sigma, \tau_j) \geqslant d(\tau_i, \tau_j)$.

There are two + two more regions in the phase diagram where the permutations which minimise the free energy can be determined by cyclically permuting the labels in the above. Most of the rest of the phase diagram can then be filled in using convexity of the free energy. That is, since f is a linear function of the x_i, if σ minimises f at two points in the phase diagram, then σ also minimises f along the segment joining these two points. This argument can only be used to fill in the whole phase diagram if the set of permutations $\Gamma(\tau_1, \tau_2, \tau_3) \coloneqq \Gamma(\tau_1, \tau_2) \cap \Gamma(\tau_2, \tau_3) \cap \Gamma(\tau_3, \tau_1)$ which simultaneously saturate the three triangle inequalities

$$d(\tau_i, \sigma) + d(\sigma, \tau_j) \geqslant d(\tau_i, \tau_j) \quad \text{for } i \neq j \,, \tag{B.4}$$

is not empty. The argument we have used to find the minima of f by considering special regions in the phase diagram and then using convexity to fill in the rest is due to [54].

From the above, we find that:

- for $x_1/x_3 + x_2/x_3 < 1$:

$$\mathcal{Z}(n) \approx d_1^{-d(\tau_1,\tau_3)} d_2^{-d(\tau_2,\tau_3)} \,. \tag{B.5}$$

  The behaviour of the sum in two other regions may be obtained by cyclically permuting the labels in the above.

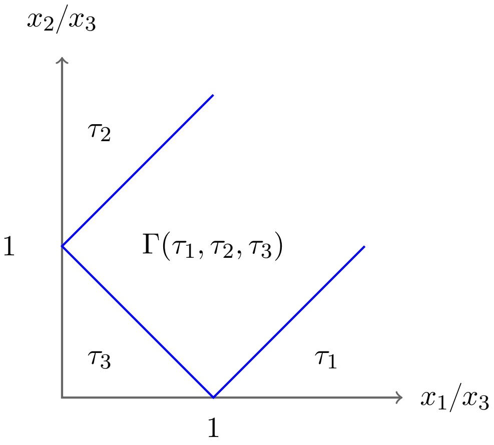

Figure 5. Phase diagram for the sum (B.1) when $\Gamma(\tau_1,\tau_2,\tau_3)$ is not empty. Along the blue lines there are more permutations which can contribute e.g. along $x_1/x_3+x_2/x_3 = 1$ the sum is dominated by the set of permuations which lie in $\Gamma(\tau_1,\tau_3) \cap \Gamma(\tau_2,\tau_3)$.

- assuming $\Gamma(\tau_1,\tau_2,\tau_3)$ is not empty, for $x_1/x_3 + x_2/x_3 > 1$, $x_2/x_1 + x_3/x_1 > 1$ and $x_3/x_2 + x_1/x_2 > 1$:

$$\mathcal{Z}(n) \approx |\Gamma(\tau_1,\tau_2,\tau_3)| \left(\frac{d_1 d_2}{d_3}\right)^{-d(\tau_1,\tau_2)/2} \left(\frac{d_2 d_3}{d_1}\right)^{-d(\tau_2,\tau_3)/2} \left(\frac{d_3 d_1}{d_2}\right)^{-d(\tau_3,\tau_1)/2} . \tag{B.6}$$

B.1 Proof

We now prove that when $\tau_N \in \{e,\eta\}$ the nested sum (3.5) in the simple model:

$$\mathcal{Z}_N(\tau_N) = \sum_{\tau_0,\ldots,\tau_{N-1}\in S_n} \prod_{p=1}^{N} d_{B_p}^{-d(\tau_{p-1},\tau_p)} d_{R_p}^{-d(\tau_{p-1},\sigma_p)} , \tag{B.7}$$

with $d_{B_p}, d_{R_p} \geqslant 1$ and $\sigma_p \in \{e,\eta\}$, is dominated by the terms with $\tau_{p-1} \in \{e,\eta\}$ for each $1 \leqslant p \leqslant N$, provided we ignore the crossover regimes. It is useful to notice that $\mathcal{Z}_N(\tau_N)$ satisfies the recursion relation

$$\mathcal{Z}_N(\tau_N) = \sum_{\tau_{N-1}\in S_n} d_{B_N}^{-d(\tau_{N-1},\tau_N)} d_{R_N}^{-d(\tau_{N-1},\sigma_N)} \mathcal{Z}_{N-1}(\tau_{N-1}) , \qquad \mathcal{Z}_0(\tau_0) = 1 . \tag{B.8}$$

First consider

$$\mathcal{Z}_1(\tau_1) = \sum_{\tau_0\in S_n} d_{B_1}^{-d(\tau_0,\tau_1)} d_{R_1}^{-d(\tau_0,\sigma_1)} . \tag{B.9}$$

This sum is of the form (B.1) so is dominated by the terms with $\tau_0 \in \{\sigma_1,\tau_1\} \subset \{e,\eta,\tau_1\}$. Using this fact we see that

$$\begin{aligned} \mathcal{Z}_2(\tau_2) &= \sum_{\tau_1\in S_n} d_{B_2}^{-d(\tau_1,\tau_2)} d_{R_2}^{-d(\tau_1,\sigma_2)} \mathcal{Z}_1(\tau_1) \\ &\approx \sum_{\tau_1\in S_n} d_{B_2}^{-d(\tau_1,\tau_2)} d_{R_2}^{-d(\tau_1,\sigma_2)} \min(d_{B_1},d_{R_1})^{-d(\tau_1,\sigma_1)} , \end{aligned} \tag{B.10}$$

is also of the form (B.1) so is dominated by the terms with $\tau_1 \in \{\sigma_1, \sigma_2, \tau_2\} \cup \Gamma(\sigma_1, \sigma_2, \tau_2) \subset \{e, \eta, \tau_2\} \cup \Gamma(e, \eta, \tau_2)$. We have assumed that $\Gamma(e, \eta, \tau_2)$ is not empty; a fact we will verify ex-post facto. Using this, (B.5) and (B.6) it is simple to show that $\mathcal{Z}_3(\tau_3)$ is also of the form (B.1) so is dominated by the terms with $\tau_2 \in \{e, \eta, \tau_3\} \cup \Gamma(e, \eta, \tau_3)$.[14] Again, we have assumed that $\Gamma(e, \eta, \tau_3)$ is not empty; a fact we will verify ex-post facto. It is not too difficult to see that this pattern continues and proceeding with the argument we find that, provided $\Gamma(e, \eta, \tau_p)$ is not empty,

$$\tau_{p-1} \in \{e, \eta, \tau_p\} \cup \Gamma(e, \eta, \tau_p) \tag{B.11}$$

for each $1 \leqslant p \leqslant N$. However, since $\tau_N \in \{e, \eta\}$, this implies that

$$\tau_{p-1} \in \{e, \eta\} \tag{B.12}$$

for each $1 \leqslant p \leqslant N$. In particular, each $\Gamma(e, \eta, \tau_p)$ is not empty, which is consistent with our assumption.

References

[1] S.W. Hawking, *Particle Creation by Black Holes*, *Commun. Math. Phys.* **43** (1975) 199 [*Erratum ibid.* **46** (1976) 206] [INSPIRE].

[2] S.W. Hawking, *Breakdown of Predictability in Gravitational Collapse*, *Phys. Rev. D* **14** (1976) 2460 [INSPIRE].

[3] I. Bena, E.J. Martinec, S.D. Mathur and N.P. Warner, *Fuzzballs and Microstate Geometries: Black-Hole Structure in String Theory*, arXiv:2204.13113 [INSPIRE].

[4] G. Penington, S.H. Shenker, D. Stanford and Z. Yang, *Replica wormholes and the black hole interior*, *JHEP* **03** (2022) 205 [arXiv:1911.11977] [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, *Entanglement Wedge Reconstruction and the Information Paradox*, *JHEP* **09** (2020) 002 [arXiv:1905.08255] [INSPIRE].

[7] 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].

[8] A. Almheiri et al., *The entropy of Hawking radiation*, *Rev. Mod. Phys.* **93** (2021) 035002 [arXiv:2006.06872] [INSPIRE].

[14] There is a slight subtlety here as the sum $\mathcal{Z}_3(\tau_3)$ can differ from (B.1) by a factor of $|\Gamma(\sigma_1, \sigma_2, \tau_2)|$. Whilst this factor depends on n, it is independent of d_{B_p} and d_{R_p} so it is reasonable to expect that we can ignore its effect if we are interested in the limit where d_{B_p} and d_{R_p} are large and eventually also the limit $n \to 1$.

[9] R. Bousso et al., *Snowmass White Paper: Quantum Aspects of Black Holes and the Emergence of Spacetime*, arXiv:2201.03096 [INSPIRE].

[10] A. Hamilton, D.N. Kabat, G. Lifschytz and D.A. Lowe, *Holographic representation of local bulk operators*, *Phys. Rev. D* **74** (2006) 066009 [hep-th/0606141] [INSPIRE].

[11] I. Heemskerk, D. Marolf, J. Polchinski and J. Sully, *Bulk and Transhorizon Measurements in AdS/CFT*, *JHEP* **10** (2012) 165 [arXiv:1201.3664] [INSPIRE].

[12] R. Bousso et al., *Null Geodesics, Local CFT Operators and AdS/CFT for Subregions*, *Phys. Rev. D* **88** (2013) 064057 [arXiv:1209.4641] [INSPIRE].

[13] B. Czech, J.L. Karczmarek, F. Nogueira and M. Van Raamsdonk, *The Gravity Dual of a Density Matrix*, *Class. Quant. Grav.* **29** (2012) 155009 [arXiv:1204.1330] [INSPIRE].

[14] A.C. Wall, *Maximin Surfaces, and the Strong Subadditivity of the Covariant Holographic Entanglement Entropy*, *Class. Quant. Grav.* **31** (2014) 225007 [arXiv:1211.3494] [INSPIRE].

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

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

[17] F. Pastawski, B. Yoshida, D. Harlow and J. Preskill, *Holographic quantum error-correcting codes: Toy models for the bulk/boundary correspondence*, *JHEP* **06** (2015) 149 [arXiv:1503.06237] [INSPIRE].

[18] X. Dong, D. Harlow and A.C. Wall, *Reconstruction of Bulk Operators within the Entanglement Wedge in Gauge-Gravity Duality*, *Phys. Rev. Lett.* **117** (2016) 021601 [arXiv:1601.05416] [INSPIRE].

[19] D. Harlow, *The Ryu-Takayanagi Formula from Quantum Error Correction*, *Commun. Math. Phys.* **354** (2017) 865 [arXiv:1607.03901] [INSPIRE].

[20] P. Hayden et al., *Holographic duality from random tensor networks*, *JHEP* **11** (2016) 009 [arXiv:1601.01694] [INSPIRE].

[21] J. Cotler et al., *Entanglement Wedge Reconstruction via Universal Recovery Channels*, *Phys. Rev. X* **9** (2019) 031011 [arXiv:1704.05839] [INSPIRE].

[22] P. Hayden and G. Penington, *Approximate Quantum Error Correction Revisited: Introducing the Alpha-Bit*, *Commun. Math. Phys.* **374** (2020) 369 [arXiv:1706.09434] [INSPIRE].

[23] C. Akers, S. Leichenauer and A. Levine, *Large Breakdowns of Entanglement Wedge Reconstruction*, *Phys. Rev. D* **100** (2019) 126006 [arXiv:1908.03975] [INSPIRE].

[24] C. Akers and G. Penington, *Leading order corrections to the quantum extremal surface prescription*, *JHEP* **04** (2021) 062 [arXiv:2008.03319] [INSPIRE].

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

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

[27] C. Akers and G. Penington, *Quantum minimal surfaces from quantum error correction*, *SciPost Phys.* **12** (2022) 157 [arXiv:2109.14618] [INSPIRE].

[28] C. Akers et al., *The black hole interior from non-isometric codes and complexity*, arXiv:2207.06536 [INSPIRE].

[29] H. Maxfield, *Bit models of replica wormholes*, arXiv:2211.04513 [INSPIRE].

[30] I.H. Kim and J. Preskill, *Complementarity and the unitarity of the black hole S-matrix*, *JHEP* **02** (2023) 233 [arXiv:2212.00194] [INSPIRE].

[31] T.J. Hollowood, S.P. Kumar, A. Legramandi and N. Talwar, *Grey-body factors, irreversibility and multiple island saddles*, *JHEP* **03** (2022) 110 [arXiv:2111.02248] [INSPIRE].

[32] P. Hayden and G. Penington, *Learning the Alpha-bits of Black Holes*, *JHEP* **12** (2019) 007 [arXiv:1807.06041] [INSPIRE].

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

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

[35] T.J. Hollowood, S.P. Kumar, A. Legramandi and N. Talwar, *Islands in the stream of Hawking radiation*, *JHEP* **11** (2021) 067 [arXiv:2104.00052] [INSPIRE].

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

[37] K. Papadodimas and S. Raju, *The unreasonable effectiveness of exponentially suppressed corrections in preserving information*, *Int. J. Mod. Phys. D* **22** (2013) 1342030 [INSPIRE].

[38] D. Stanford, *More quantum noise from wormholes*, arXiv:2008.08570 [INSPIRE].

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

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

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

[42] V.E. Hubeny, M. Rangamani and E. Tonni, *Global properties of causal wedges in asymptotically AdS spacetimes*, *JHEP* **10** (2013) 059 [arXiv:1306.4324] [INSPIRE].

[43] T.J. Hollowood, S. Prem Kumar and A. Legramandi, *Hawking radiation correlations of evaporating black holes in JT gravity*, *J. Phys. A* **53** (2020) 475401 [arXiv:2007.04877] [INSPIRE].

[44] Z. Gyongyosi et al., *Black Hole Information Recovery in JT Gravity*, arXiv:2209.11774 [DOI:10.1007/JHEP01(2023)139] [INSPIRE].

[45] S. Goldstein, J.L. Lebowitz, R. Tumulka and N. Zanghi, *Canonical Typicality*, *Phys. Rev. Lett.* **96** (2006) 050403 [cond-mat/0511091] [INSPIRE].

[46] H. Liu and S. Vardhan, *Entanglement entropies of equilibrated pure states in quantum many-body systems and gravity*, *PRX Quantum* **2** (2021) 010344 [arXiv:2008.01089] [INSPIRE].

[47] D. Marolf and H. Maxfield, *Observations of Hawking radiation: the Page curve and baby universes*, *JHEP* **04** (2021) 272 [arXiv:2010.06602] [INSPIRE].

[48] T.J. Hollowood and N. Talwar, to appear.

[49] X.-L. Qi, *Entanglement island, miracle operators and the firewall*, *JHEP* **01** (2022) 085 [arXiv:2105.06579] [INSPIRE].

[50] Y. Chen, *Pulling Out the Island with Modular Flow*, *JHEP* **03** (2020) 033 [arXiv:1912.02210] [INSPIRE].

[51] L. Susskind, *The Typical-State Paradox: Diagnosing Horizons with Complexity*, *Fortsch. Phys.* **64** (2016) 84 [arXiv:1507.02287] [INSPIRE].

[52] S.D. Mathur, *The information paradox: conflicts and resolutions*, *Pramana* **79** (2012) 1059 [arXiv:1201.2079] [INSPIRE].

[53] S.D. Mathur and D. Turton, *Comments on black holes I: The possibility of complementarity*, *JHEP* **01** (2014) 034 [arXiv:1208.2005] [INSPIRE].

[54] C. Akers, T. Faulkner, S. Lin and P. Rath, *Reflected entropy in random tensor networks*, *JHEP* **05** (2022) 162 [arXiv:2112.09122] [INSPIRE].

[55] B. Collins, *Moments and Cumulants of Polynomial random variables on unitary groups, the Itzykson-Zuber integral and free probability*, *Int. Math. Res. Not.* **2003** (2003) 953 [math-ph/0205010] [DOI:10.48550/arXiv.math-ph/0205010].

[56] B. Collins and P. Śniady, *Integration with Respect to the Haar Measure on Unitary, Orthogonal and Symplectic Group*, *Commun. Math. Phys.* **264** (2006) 773 [math-ph/0402073].

[57] J. Kudler-Flam, V. Narovlansky and S. Ryu, *Distinguishing Random and Black Hole Microstates*, *PRX Quantum* **2** (2021) 040340 [arXiv:2108.00011] [INSPIRE].

[58] J. Kudler-Flam and Y. Kusuki, *On quantum information before the Page time*, *JHEP* **05** (2023) 078 [arXiv:2212.06839] [INSPIRE].