Signatures of global symmetry violation in relative entropies and replica wormholes

Yiming Chen and Henry W. Lin

Jadwin Hall, Princeton University,
Princeton, NJ 08540, U.S.A.

E-mail: ymchen.phys@gmail.com, hwlin@princeton.edu

ABSTRACT: It is widely believed that exact global symmetries do not exist in theories that admit quantum black holes. Here we propose a way to quantify the degree of global symmetry violation in the Hawking radiation of a black hole by using certain relative entropies. While the violations of global symmetry that we consider are non-perturbative effects, they nevertheless give $\mathcal{O}(1)$ contributions to the relative entropy after the Page time. Furthermore, using "island" formulas, these relative entropies can be computed within semi-classical gravity, which we demonstrate with explicit examples. These formulas give a rather precise operational sense to the statement that a global charge thrown into an old black hole will be lost after a scrambling time.

The relative entropies considered here may also be computed using a replica trick. At integer replica index, the global symmetry violating effects manifest themselves as charge flowing through the replica wormhole.

Contents

1 Introduction

An old conjecture states that in any theory of quantum gravity with black holes, there are no exact global symmetries [1–6]. A semiclassical treatment of black hole evaporation [7] suggests that the Hawking radiation has a thermal spectrum, independent of whatever the global charge of the matter that formed the black hole. After the black hole evaporates completely, any initial global charge is lost (up to fluctuations in the thermal ensemble). One might hope that the charge can be stored in some remnant of the black hole, but such remnants would violate entropy bounds [4] among other problems [8]. These arguments are most clear at the final stage of the evaporation, where the black hole is not big enough to store the information about the infalling charge, while they are less clear when the black hole is still large, say at the Page time of the black hole evaporation. Said differently, an experimentalist who wishes to check whether symmetry violation has occurred must gather almost all of the Hawking radiation from the black hole. If she is unable to capture the

high energy modes at the endpoint of evaporation, or if she is unable to access certain fields (e.g. dark matter fields), she may not be able to conclusively convince herself that charge conservation is violated in nature.

Since we expect that the global symmetry violation happens throughout the evaporating process, it is therefore natural to ask that whether there are some quantities that can be computed with only some portion of the Hawking radiation, that are able to show explicitly the global symmetry has been violated. (There is a strong analogy here with the Page curve [9]: unitarity implies that the entropy of *all* the Hawking radiation from a fully evaporated black hole is zero; Page famously showed that unitarity also has implications for a large subset of the radiation.) One obvious candidate is to compute correlators which would be zero if the global symmetry is exact. However, if there are no explicit symmetry violating processes in the semiclassical physics, these correlators are expected to have non-perturbatively small values $\mathcal{O}(e^{-S_{\mathrm{BH}}})$, where S_{BH} is the Bekenstein-Hawking entropy of the black hole. Computing these quantities will require non-perturbative knowledge beyond Hawking's description. To be clear, we are not claiming that the approximate global symmetry, such as the $B - L$ symmetry in nature, cannot be violated within the semiclassical physics. What we are imagining is the worst case scenario, namely even if the symmetry is preserved semiclassically, non-perturbative gravitational effect will eventually violate it.

Although understanding how to compute such exponentially small violations is a worthy goal for the future, the point of this paper is to point out there exist other quantities which quantify the symmetry violation in the Hawking radiation, that can be computed just with the semiclassical knowledge, and give $\mathcal{O}(1)$ results. Our discussion is motivated by the recent progress in deriving the unitary Page curve [9] for an evaporating black hole within the semiclassical description [10, 11]. In those developments, the key tool is the gravitational fine grained entropy formula, or the quantum extremal surface (QES) prescription [12–15]. When applying the formula to compute the von Neumann entropy of the Hawking radiation after the Page time, one finds an island in the black hole interior that belongs to the entanglement wedge of the Hawking radiation [16]. In this paper, we consider the relative entropies between different density matrices of the Hawking radiation R. More precisely, we consider the exact density matrix of the Hawking radiation $\rho_1 = \rho_{\mathrm{exact}}(R)$, and the density matrix $\rho_2 = U_R(g)\rho_{\mathrm{exact}}(R)U_R^\dagger(g)$ after applying a global symmetry transformation $U_R(g), g \in G$ to the radiation. If we form the black hole from neutral (singlet) matter, then if the global symmetry were exact, we would expect $\rho_2 = \rho_1$ since a reduced density matrix of the system should be invariant under the symmetry transformation. Thus a non-zero relative entropy $S(\rho_2|\rho_1) > 0$ is a way to quantify the degree to which the global symmetry is violated. Similar to the von Neumann entropy, the relative entropy can be computed with only the knowledge of the semiclassical physics [17], and we find it to be an $\mathcal{O}(1)$ result whenever an island arises. We demonstrate the calculation explicitly using simple examples. The relative entropy that we compute quantifies the amount of global symmetry violation in the Hawking radiation, and we discuss how it depends on various parameters of the problem. As we will see in section 2, the relative entropy is zero before the Page time, but becomes non-zero at the Page time and grows as more radiation comes out.

The island in the QES prescription is closely related to the replica wormhole geometry in the replica trick for deriving the entropy using the gravitational path integral [18–20]. It is thus natural to ask what roles they play in showing that the global symmetry is violated. In section 3 we discuss the replica versions of the relative entropy, and point out that they are non-zero exactly due to charge flowing through the replica wormhole. This resonates with a long history of works anticipating that Euclidean wormholes violate global symmetries [2, 21]. In previous arguments, it was never completely clear whether wormholes should be included or not in a particular calculation; here, unitarity demands them to be included. We contrast the global symmetry case with a gauge symmetry, which is allowed because charge cannot flow through the wormhole due to the gauge constraint (Gauss's law). An exact global symmetry of a holographic theory must therefore be realized as a gauge symmetry in gravity.

We also consider in section 2.3 a setup where the bulk relative entropy may be enhanced by creating pairs of particles with opposite charges in the bath, and waiting a scrambling time for the particles to fall into the island. This is related to the Hayden-Preskill [22] experiment, as we elaborate on in section 4 along with other discussions and conclusions.

It is worth noting that a recent paper by Harlow and Shaghoulian [23] also connected recent developments in the black hole information problem with the violation of global symmetries, using arguments along the lines of [5, 6]. Our result is based on similar assumptions as theirs, but in some sense we take one step further by quantifying and computing the amount that an approximate global symmetry is violated by non-perturbative effects.

Note added. as we were finishing the manuscript, we learned that [24] are considering similar issues in an upcoming work.

2 Quantifying global symmetry violation with relative entropy

2.1 General argument

We start by reviewing the QES prescription and the relative entropy formula. In the context of an evaporating black hole, the QES prescription as applied to the exact density matrix $\rho_{\text{exact}}(R)$ of the Hawking radiation states

$$S\left(\rho_{\text{exact}}(R)\right) = \text{Min}\left\{\text{Ext}_I\left[\frac{A(\partial I)}{4G_N} + S(\rho_{\text{semi}}(R \cup I))\right]\right\}, \tag{2.1}$$

where on the right hand side, one extremizes over the generalized entropy function over possible islands in the region where gravity is dynamical, and choose the minimal one if there are multiple extremums. Note that the right hand side of (2.1) is calculable within the semiclassical approximation. The region $R \cup I$ on the right hand side is called the entanglement wedge [25–27] of the region R. As first recognized in the context of the AdS/CFT correspondence [17], the validity of (2.1) among a family of "nearby" states implies an equality between the relative entropies in the exact description and the semi-classical description:

$$S\left(\rho_{\text{exact}}(R)|\sigma_{\text{exact}}(R)\right) = S\left(\rho_{\text{semi}}(R \cup I)|\sigma_{\text{semi}}(R \cup I)\right), \tag{2.2}$$

where the relative entropy $S(\rho|\sigma)$ is defined as

$$S(\rho|\sigma) \equiv \mathrm{tr}\left[\rho(\log\rho - \log\sigma)\right], \tag{2.3}$$

which is a non-negative quantity quantifying how different ρ and σ are. The relative entropy formula has further subleading corrections, but in this paper we will only be focusing on the leading order result as in (2.2).

Our argument will give non-trivial results in cases where there is an island in the entanglement wedge of the radiation. The phenomena of islands has been proposed for a variety of scenarios involving different dimensions and also different signs of the cosmological constant (see [28–43] for some works in this direction and more in a recent review [44]). Our argument is general, but we will use the toy examples in [28] as a concrete set-up for our discussion. More specifically, we will consider the case with a zero temperature black hole in the two-dimensional Jackiw-Teitelboim (JT) gravity [45–47] coupled to a flat space bath. There is a two-dimensional CFT with central charge $c \gg 1$ propagating both in the JT gravity region and in the flat space region. This gravity description is semiclassical, namely we will be neglecting nonperturbative effects that are suppressed by e^{-S_0}, where S_0 is the extremal entropy of the black hole. We will also assume that the model has a dual nonperturbative description as a quantum mechanical dot coupled to a half infinite line where the CFT lives. The two descriptions of the system are illustrated in figure 1(a). For the ground state of the system, the metric and the dilaton profile in the AdS$_2$ region are

$$ds^2 = \frac{-dt^2 + dx^2}{x^2}, \quad x < 0, \quad \phi = S_0 - \frac{\phi_r}{x}, \tag{2.4}$$

where ϕ_r defines a length scale of the problem. The metric in the flat space region is given by

$$ds^2 = -dt^2 + dx^2, \quad x > 0, \tag{2.5}$$

and the conformal fields are in the Minkowski vacuum with respect to the global time coordinate t. For a region R in the flat space region that is large enough, applying the QES prescription to compute the entropy (and the location of the entanglement wedge), one finds that the entanglement wedge of R includes an island I in the JT gravity region (see figure 1(b)).

We consider situations in which there is a global symmetry in the semiclassical gravity description, with a symmetry group G that can generally be discrete or continuous, and possibly non-Abelian. We will assume that the symmetry is *not* gauged in the gravity description, leaving a discussion of gauge symmetries to section 3.3. In terms of the exact "boundary" description, the flat space CFT has an exact global symmetry without coupling to the boundary quantum mechanical system, and the coupling breaks this symmetry explicitly. However, the couplings that break the symmetry should be nonperturbatively small so that the global symmetry is still realized in the gravity region in the semiclassical description.

To simplify the discussion, we will first consider the case where the entire system is in a state which is invariant under the global symmetry transformation in the semiclassical
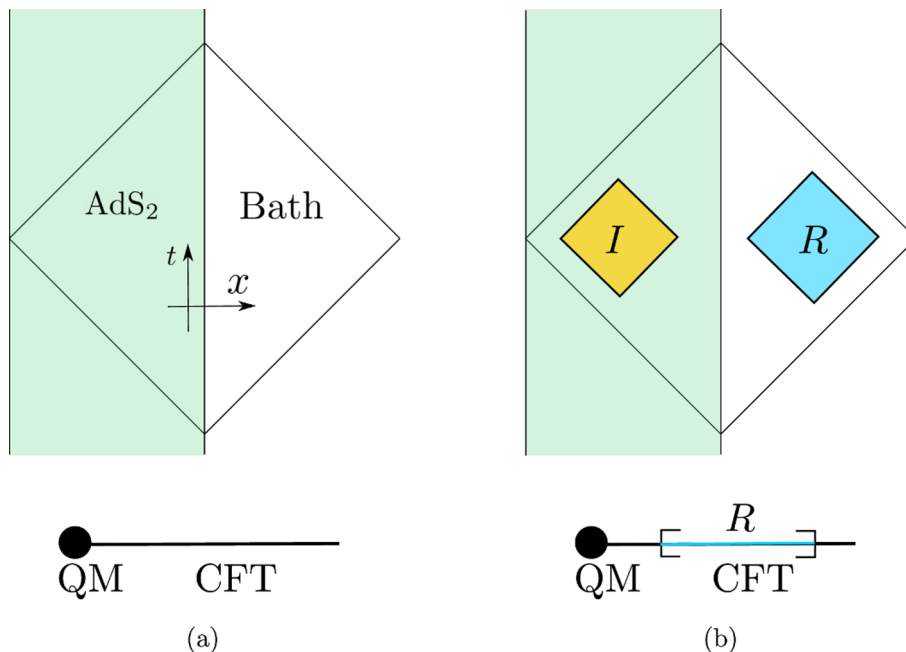
Figure 1. (a) A zero temperature black hole in JT gravity coupled to bath. It has a nonperturbative description in terms of a boundary quantum mechanical system coupled to a CFT in half infinite space. (b) When we compute the entropy of a region R in the bath that is large enough, there will be an island I in the gravity region.

description. For example, we could consider the ground state of the model in figure 1(a), for which the semiclassical state has the global symmetry.[1] If the state of the entire system is invariant under a symmetry transformation U(g), then when we restrict to a spatial subregion R, the reduced density matrix $\rho(R)$ should also be invariant under the symmetry transformation restricted in R, which we denote by $U_R(g)$. (We will mostly leave the dependence on the group element $g \in G$ implicit from now on.) The reason is that the symmetry transformation U of a global symmetry has the split property [5]; morally, this means it can be written as a product of the symmetry transformation in R and its compliment $\bar{R}$, $U = U_R U_{\bar{R}}$.[2] It follows that

$$U_R \rho_R U_R^\dagger = \text{tr}_{\bar{R}} \left(U_R \rho U_R^\dagger \right) = \text{tr}_{\bar{R}} \left(U_{\bar{R}} U_R \rho U_R^\dagger U_{\bar{R}}^\dagger \right) = \text{tr}_{\bar{R}} \left(U \rho U^\dagger \right) = \rho_R. \qquad (2.6)$$

Applied to our case, since the semiclassical state has the symmetry, we have

$$U_R \rho_{\text{semi}}(R) U_R^\dagger = \rho_{\text{semi}}(R), \qquad (2.7)$$

or equivalently

$$S \left(U_R \rho_{\text{semi}}(R) U_R^\dagger \big| \rho_{\text{semi}}(R) \right) = 0. \qquad (2.8)$$

---

[1]We will be assuming that the semiclassical global symmetry is not spontaneously broken throughout the paper.

[2]In a continuum quantum field theory, there is an edge term U_{edge} that depends on how we regulate the boundary between R and $\bar{R}$. Since it is not essential to our discussion, we will keep it implicit.

This motivates us to compare the two density matrices $\rho_{\text{exact}}(R)$ and $U_R^\dagger \rho_{\text{exact}}(R) U_R$, by computing the relative entropy between them, i.e. $S\left(U_R^\dagger \rho_{\text{exact}}(R) U_R \middle| \rho_{\text{exact}}(R)\right)$. Interestingly, to compute this quantity, we do not need to have a complete knowledge about $\rho_{\text{exact}}(R)$. We can compute it just from the semiclassical description, by applying the relative entropy formula (2.2):[3]

$$S\left(U_R \rho_{\text{exact}}(R) U_R^\dagger \middle| \rho_{\text{exact}}(R)\right) = S\left(U_R \rho_{\text{semi}}(R \cup I) U_R^\dagger \middle| \rho_{\text{semi}}(R \cup I)\right). \qquad (2.9)$$

We should stress that the transformation U_R on the left hand side is a completely well defined operator in the nonperturbative description. For a continuous symmetry with Noether current J_a^μ, it is simply given by $e^{i\epsilon_a \int_R *J_a}$. An important assumption in the formula however is that when we pass from the exact description to the semiclassical description in (2.9), the transformation only acts in the region R but not the island, and still acts as the same symmetry transformation in the semiclassical theory.[4] One should be cautious about this, because the entanglement wedge of R includes both the region itself and the island, and in principle a general operation in R in the exact description can act both in the region R and the island I. Indeed, as advocated in [48], a modular flow using the exact modular Hamiltonian of R corresponds to a modular flow on $R \cup I$ with the semiclassical modular Hamiltonian. However, in contrast with the exact modular Hamiltonian, which is expected to be non-local and encodes the fine grained information of the state, the operator U_R factorizes into local pieces and does not contain information about the state. In other words, it is a "simple" operation. A more careful argument is the following. We can divide the region R into smaller pieces $\{R_i\}$, and the operator U_R factorizes as

$$U_R = \prod_i U_{R_i}. \qquad (2.10)$$

If we divide R such that R_i are small enough, the entanglement wedge of each R_i will only contain the region itself. Thus each U_{R_i} should only act within R_i without any ambiguity, so the product should only act in the union of R_i, which is R. This is part of the idea that has been used in [5, 6, 23] to argue that there cannot be an exact global symmetry. In section 3, we will also justify the formula (2.9) from the replica computation point of view.

After justifying the formula (2.9), we now turn to its implications. The density matrix $\rho_{\text{semi}}(R \cup I)$ is invariant under the semiclassical global symmetry transformation on the union of R and I, but it is *not* invariant under U_R alone. This is because the fields in R are entangled with the fields in I and so $\rho_{\text{semi}}(R \cup I) \neq \rho_{\text{semi}}(R) \otimes \rho_{\text{semi}}(I)$; otherwise it would be invariant under U_R because of (2.7). For this reason, for a fixed U_R, one expects the result to be larger if the region R and the island have larger mutual information. To summarize, we expect that the bulk formula (2.9) gives a nonzero relative entropy when there is an island:

$$S\left(U_R \rho_{\text{exact}}(R) U_R^\dagger \middle| \rho_{\text{exact}}(R)\right) = S\left(U_R \rho_{\text{semi}}(R \cup I) U_R^\dagger \middle| \rho_{\text{semi}}(R \cup I)\right) = \mathcal{O}(1), \qquad (2.11)$$

---

[3]The relative entropy is valid either by considering a transformation U_R that is close to identity in the case of continuous symmetry, or by considering a global symmetry that only acts on a small subset of fields (note that we have $c \gg 1$).

[4]We thank Ahmed Almheiri and Edgar Shaghoulian for asking about this assumption.

where $\mathcal{O}(1)$ denotes that this quantity is not suppressed by e^{-S_0}. As a side note, assuming that the global semiclassical state is invariant under the symmetry, $\rho_{\text{semi}}(R \cup I)$ is invariant under $U_R U_I$, so we can also compute the relative entropy via

$$S\left(U_R \rho_{\text{exact}}(R) U_R^\dagger \middle| \rho_{\text{exact}}(R)\right) = S\left(U_I^\dagger \rho_{\text{semi}}(R \cup I) U_I \middle| \rho_{\text{semi}}(R \cup I)\right). \qquad (2.12)$$

The quantity $S\left(U_R \rho_{\text{semi}}(R \cup I) U_R^\dagger \middle| \rho_{\text{semi}}(R \cup I)\right)$ is computable solely using the semi-classical description. Since the von Neumann entropies of $U_R \rho_{\text{semi}}(R \cup I) U_R^\dagger$ and $\rho_{\text{semi}}(R \cup I)$ are the same as they only differ by an unitary, the relative entropy can be expressed as the difference of the expectation value of the modular Hamiltonian $K_{\text{semi}}(R \cup I) \equiv -\log \rho_{\text{semi}}(R \cup I)$:

$$S\left(U_R \rho_{\text{semi}}(R \cup I) U_R^\dagger \middle| \rho_{\text{semi}}(R \cup I)\right) = \Delta\langle K_{\text{semi}}(R \cup I)\rangle, \qquad (2.13)$$

where we take the expectation value of the modular Hamiltonian in the state after acting with U_R, and subtract that in the original state. In simple cases, this computation can be carried out explicitly, as we will demonstrate in section 2.2.

Before we move on to concrete examples, let us offer some qualitative comments. For the relative entropy to be nonzero, we need the region R to be large enough so that R claims the island. If we divide R into small pieces, then when we only look at the relative entropies for the small spatial pieces, we will not see such an effect.

Similar statement holds if we divide the fields in R into small subsets. For example, imagine the CFT is N copies of free fermions ($N \gg 1$), which has an U(N) global symmetry. We could consider a symmetry transformation U_R that only acts among a small subset of the fermions, say M fermions with $M \sim \mathcal{O}(1)$. If we only look at the reduced density matrix of the M fermions while tracing out the rest, since it does not claim the island, we would conclude that it is invariant under the symmetry transformation. However, the density matrix of all the N fermions will not be invariant by (2.11). This suggests that the main contribution to the order one result in (2.11) comes from correlations between the M fermions and the rest of the fermions. For this to be true, there must be sufficient correlation between the M fermions and the rest of the fields at the non-perturbative level, despite the fact that they are decoupled and uncorrelated in the semiclassical description.[5] This correlation can also be quantified using the gravitational fine grained entropy formula, and we comment more on this in appendix A.

2.2 Example: U(1) global symmetry of Dirac fermions

As a concrete, calculable example, we consider a setup where the CFT is a tensor product of a massless Dirac fermion and some other CFT with central charge $c - 1$ which we will not specify. For simplicity, we will assume that in the semiclassical description, this fermion field does not couple to the rest of the CFT directly, and we also neglect the Schwarzian fluctuations. In the semiclassical description, the model has a U(1) global

---

[5] Here we are neglecting the universal coupling through the boundary graviton in the semiclassical description. In the current model, this can be justified in the limit of $c \gg 1$ and S_0/c not too large.

symmetry which rotates the fermion as $\psi \to e^{i\theta}\psi$. The time component of the conserved current corresponding to this symmetry is the fermion number density:

$$J_0 \equiv \psi_+^\dagger \psi_+ + \psi_-^\dagger \psi_-, \qquad (2.14)$$

and the transformation we apply in R is

$$U_R \equiv \exp\left(i\alpha Q_R\right) = \exp\left(i\alpha \int_R dx\, J_0\right). \qquad (2.15)$$

As we explained in (2.13), the relative entropy is given by the change of the expectation value of the semiclassical modular Hamiltonian $K_{\text{semi}}(R \cup I)$:

$$S\left(U_R \rho_{\text{exact}}(R) U_R^\dagger \middle| \rho_{\text{exact}}(R)\right) = \langle 0| U_R^\dagger K_{\text{semi}}(R \cup I) U_R |0\rangle - \langle 0| K_{\text{semi}}(R \cup I) |0\rangle, \qquad (2.16)$$

where the $|0\rangle$ state is the vacuum state of the fermion in the semiclassical description. Since the fermion is massless, we can neglect the warp factor inside the AdS region. In other words, up to local terms at the boundary of the island that depends on the warp factor, the modular Hamiltonian $K_{\text{semi}}(R \cup I)$ is just the modular Hamiltonian for two disjoint intervals in flat space, whose form was given explicitly in [49] which we review in appendix. B. We will not need the detailed form of the modular Hamiltonian here to understand the calculation, but only need its basic structure. For the vacuum state, the modular Hamiltonian is quadratic in the fermion operators. It can be separated into two pieces, the local piece K_{loc} which couples operators within R or I, and the non-local piece K_{noloc} that couples a fermion operator in R to another in I. It is not hard to see that

$$U_R^\dagger K_{\text{loc}} U_R = K_{\text{loc}}, \qquad (2.17)$$

as K_{loc} can be expressed as an integral of the local energy density operator $T(x)$, which is invariant under U_R (the piece in the island is invariant trivially because it is spacelike separated from U_R). On the other hand, K_{noloc} transforms non-trivially under U_R. K_{noloc} has the form

$$K_{\text{noloc},\pm} = i \int_I dx \int_R dy\, K(x,y)(\psi_{\pm,I}^\dagger(x)\psi_{\pm,R}(y) - \psi_{\pm,R}^\dagger(y)\psi_{\pm,I}(x)), \qquad (2.18)$$

where the kernel $K(x,y)$ is given explicitly in appendix B. The symmetries act as

$$U_R^\dagger \psi_{\pm,I}^\dagger(x)\psi_{\pm,R}(y) U_R = e^{i\alpha}\psi_{\pm,I}^\dagger(x)\psi_{\pm,R}(y), \quad U_R^\dagger \psi_{\pm,R}^\dagger(y)\psi_{\pm,I}(x) U_R = e^{-i\alpha}\psi_{\pm,R}^\dagger(y)\psi_{\pm,I}(x). \qquad (2.19)$$

By noting that $\langle 0| \psi_{\pm,I}^\dagger(x)\psi_{\pm,R}(y) |0\rangle = -\langle 0| \psi_{\pm,R}^\dagger(y)\psi_{\pm,I}(x) |0\rangle$, we obtain

$$S\left(U_R \rho_{\text{exact}}(R) U_R^\dagger \middle| \rho_{\text{exact}}(R)\right) = (\cos\alpha - 1)\langle 0| K_{\text{noloc}} |0\rangle. \qquad (2.20)$$

The expression is periodic in 2π since the transformation with $\alpha/2\pi = n \in Z$ acts trivially. The expectation value of K_{noloc} is computed explicitly in appendix B. In Poincaré coordinates x, if we parametrize the regions as $I = (a_1, b_1), R = (a_2, b_2)$, then

$$S\left(U_R \rho_{\text{exact}}(R) U_R^\dagger \middle| \rho_{\text{exact}}(R)\right) = \sin^2\frac{\alpha}{2}\left(1 + \frac{(2\eta-1)\arctan\sqrt{\frac{\eta}{1-\eta}}}{\sqrt{\eta(1-\eta)}}\right), \quad \eta \equiv \frac{(b_1-a_1)(b_2-a_2)}{(a_2-a_1)(b_2-b_1)}. \qquad (2.21)$$

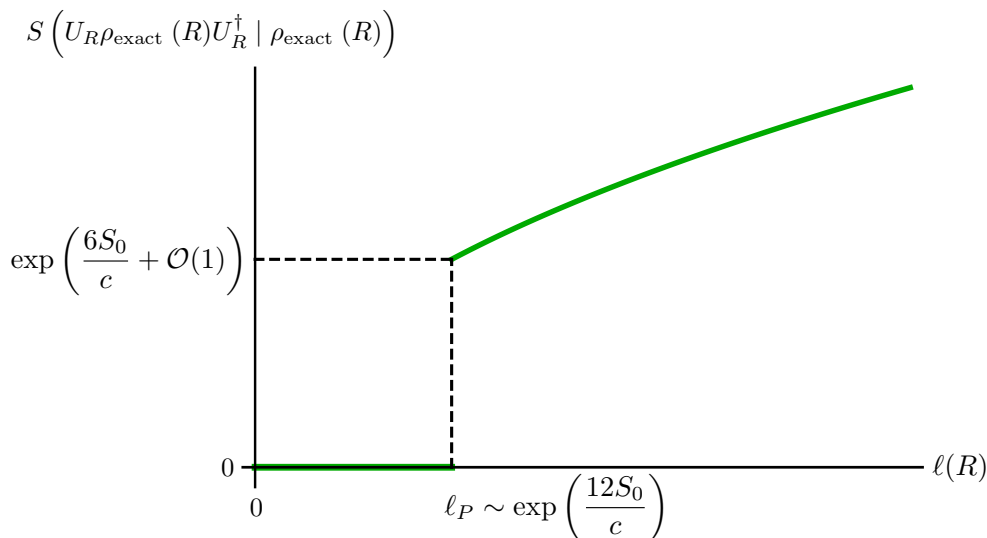**Figure 2**. The relative entropy in the free fermion model, as a function of the size of the region R at fixed non-zero α. The horizontal axis is $\ell(R)$ the length of R in units of ϕ_r/c ($\ell(R) \equiv c(b_2 - a_2)/\phi_r \approx cb_2/\phi_r$). In the semi-classical approximation, there is a discontinuity in the relative entropy at the "Page length" $\ell_P(R) \sim e^{12S_0/c}$ when an island appears in the gravity region, see (2.24). It might seem that the relative entropy is large because it has S_0 in the exponential, but recall that in this model the ratio S_0/c is kept finite and not very large [28], so the relative entropy is $\mathcal{O}(1)$.

We see that (2.21) diverges as $1/\sqrt{1-\eta}$ when $\eta \to 1$, where the island and the bath region get closer. This is consistent with the intuition that the result should be larger if there is more mutual information between R and I.[6]

As discussed in [28], if we choose the bath region

$$a_2 \ll \frac{\phi_r}{c} \ll b_2, \quad \log\left(\frac{cb_2}{\phi_r}\right) > \frac{12S_0}{c} + \mathcal{O}(1), \tag{2.22}$$

then the island will be at

$$a_1 \approx -b_2, \quad b_1 \approx -\frac{6\phi_r}{c}. \tag{2.23}$$

Plugging (2.22) and (2.23) into (2.21), we find

$$S\left(U_R\rho_{\text{exact}}(R)U_R^\dagger \middle| \rho_{\text{exact}}(R)\right) \approx \frac{\pi}{4}\sin^2\frac{\alpha}{2}\sqrt{\frac{cb_2}{3\phi_r}}. \tag{2.24}$$

We observe that the relative entropy grows with the total central charge c, even though we are considering the symmetry that only involves one fermion. This confirms our previous expectation that the symmetry violation is contained in the correlation between the fermion field and other fields. We stress that $\rho_{\text{exact}}(R)$ is the density matrix for all the fields, not just the one fermion field. Note that ϕ_r/c in (2.24) is roughly the time it takes for a particle to reach the island if it is released from the boundary. We will give a physical explanation of why this "island time" is relevant in section 2.3. We sketch the result (2.24) in figure 2.

---

[6]However, it is interesting to note that it grows faster than the mutual information as $\eta \to 1$, since the mutual information only grows logarithmically.

There is an analogue of the "Page time" (perhaps more aptly called the Page length) ℓ_P in this problem that is the minimal size of R to claim the island. Below the minimal size, the relative entropy is zero, while above that size, the relative entropy becomes finite and grows with the size. At finite S_0, we expect to have a smooth curve instead of a sharp transition, see [50–52].

2.3   Increasing the relative entropy with charged particles

So far, we've only considered the ground state of the model in figure 1. It is natural to ask whether we can choose different bulk semi-classical states which will have a larger relative entropy when there is an island. To motivate the discussion, let us first look at a rough description of the semi-classical vacuum state. There are three regions, the island I, the radiation R, and the region in between — call it S for sea. The global state is charge neutral but locally there can be vacuum fluctuations of the charge density. We will focus on two possible forms of the fluctuation, and look at the density matrices they lead to. There are of course contributions from other types of terms and cross terms, but our goal here is just to draw some basic intuition. The first case corresponds to virtual loops that run between the sea and the radiation, as represented by ℓ_2 in figure 3(a):

$$|\text{vac}\rangle_{\text{bulk}} \sim \sum_q |0\rangle_I \, |-q\rangle_S \, |q\rangle_R \,, \tag{2.25}$$

where $Q_i \, |q\rangle_i = q_i \, |q\rangle_i$, $i = I, S, R$. The density matrix that it leads to is

$$\rho_{\text{semi}}(R \cup I) \sim \sum_q \left(|0\rangle_I \, \langle 0|_I \otimes |q\rangle_R \, \langle q|_R \right). \tag{2.26}$$

It is invariant under U_R, and thus will not contribute to the relative entropy. The second case involves loops that run between the island and the radiation (loop ℓ_1 in figure 3(a)):

$$|\text{vac}\rangle_{\text{bulk}} \sim \sum_q |-q\rangle_I \, |0\rangle_S \, |q\rangle_R \,, \tag{2.27}$$

and it leads to

$$\rho_{\text{semi}}(R \cup I) \sim \sum_{q,\tilde{q}} \left(|-q\rangle_I \, \langle -\tilde{q}|_I \otimes |q\rangle_R \, \langle \tilde{q}|_R \right). \tag{2.28}$$

We see that the semiclassical density matrix is not invariant under U_R, and will lead to a non-trivial dependence on α in the relative entropy via (2.2). These loops are also the virtual processes that build up the entanglement between the island and the radiation, so we see that the relative entropy is closely related to entanglement, or mutual information. See appendix A for some further discussion.

The above toy model is supposed to depict the quantum fluctuations in the vacuum. However, it also suggests a way to increase the bulk relative entropy by making on shell particles with correlated charges. To neglect the fluctuations that already exist in the vacuum state, we could consider a matter theory which consists of a large c CFT + a massive "probe" theory with an U(1) symmetry (for example, a single massive free fermion field).
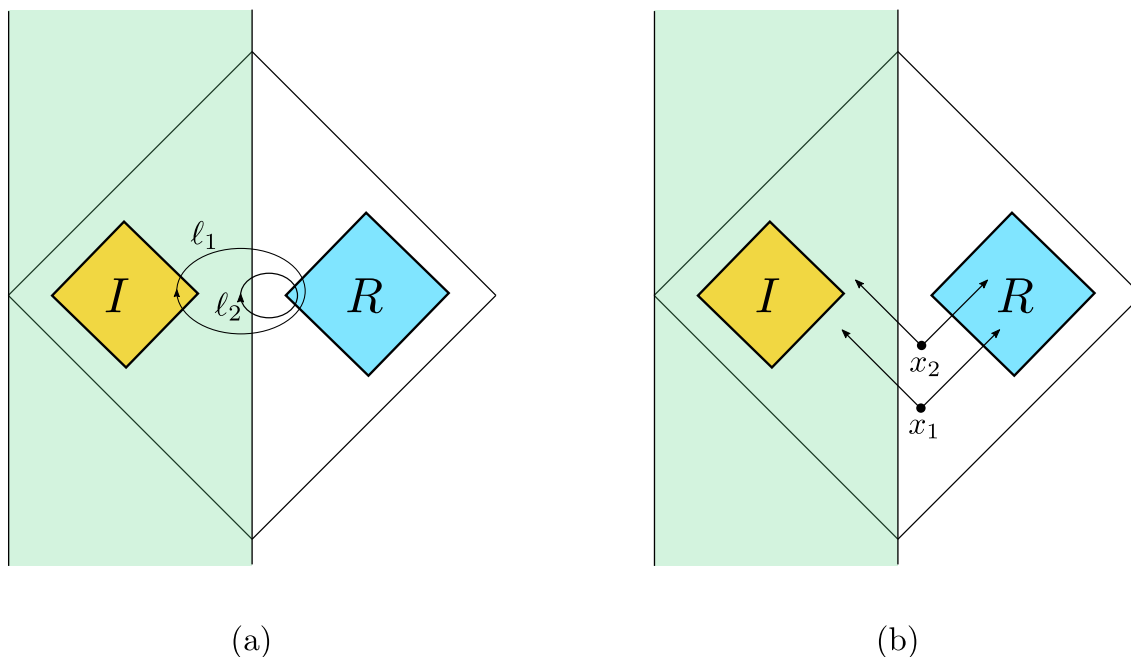
Figure 3. (a) We consider two virtual processes ℓ_1 and ℓ_2 where charged particles run in loops. Only the larger loop ℓ_1 contributes to the bulk relative entropy (as well as the entanglement between I and R). (b) We imagine the decay of a neutral particle into 2 fermions at x_1 and x_2. Only the decay at x_1 contributes significantly to the bulk relative entropy. Note also that in an evaporating black hole, a morally similar picture would also predict that the relative entropy increases after the Page time because each Hawking mode will be entangled with a mode of opposite charge behind the horizon.

By "probe" field, we mean that when we can neglect the massive theory in determining the location of the island. When we compute the relative entropy with U(1) symmetry generators acting on the massive theory using (2.13), we expect it to be suppressed by $e^{-m\ell}$ where ℓ is the proper distance between the island and the bath region R, and is thus small for large mass. So if we create many pairs of particles, the leading contribution to the relative entropy will come from the real particles instead of vacuum fluctuations. In other words, we can interpret the charge fluctuations in (2.25) and (2.27) as coming from real particles.

Imagine that we start with a massive neutral particle at rest in the bath region. It then decays into two massive free fermions, with charge $+q$ and $-q$. These fermions propagate in opposite directions: one towards the black hole, and the other towards null infinity. However, we do not know which charge went where. It is interesting to consider the dependence of the relative entropy on the location x of the fermion. If the particle which falls into the black hole is not in the island, then we have a state similar to (2.25), and it will not increase the relative entropy significantly. However, if we drop in the particle and then wait some time t_{island}, the particle will appear in the island, see figure 3. In cases where the island is near the horizon, for example in the evaporating black hole, or in the 2-sided setup at late times, we will need to wait about a scrambling time for a perturbation applied

near the boundary of the gravity region.[7] This will lead to a nonzero relative entropy, by considerations nearly identical to those in (2.27) and (2.28).

We would like to interpret the nonzero relative entropy in this case as saying that the charge is lost from the outside observer point of view. If the symmetry was exact, we should find a zero relative entropy at all time. The fact that we have a nonzero relative entropy after creating the particle pair simply means that the charge that is outside the region R is not conserved. From our discussion, we see that this non-conservation happens at the time scale which is the scrambling time. We will return to this point in the discussion.

2.4 Evaporating black hole

In the discussion of section 2.1 and 2.2, we mainly considered time independent cases where a black hole is in thermal equilibrium with the environment. However, our discussion also applies to general time dependent cases, such as evaporating black holes. Consider a theory that has a global symmetry in the semiclassical description, and we start from a pure state that has zero global charge and form a black hole. If the symmetry is not violated nonperturbatively, then if we look at the Hawking radiation emitted from the black hole, its density matrix should always be block diagonal in the charge basis, and thus is invariant under the symmetry transformation acting on the Hawking radiation. However, after the Page time, since the entanglement wedge of the Hawking radiation contains an island in the black hole interior, we can similarly compute relative entropies and see that the global symmetry is violated from a nonzero result. We also expect the relative entropy to grow as one collects more radiation from the black hole.

Of course, it is well known how to see that global symmetry is violated by looking at the final stage of evaporation as we reviewed in the introduction [4]. Here we are relating the violations of global symmetry to the appearance of an "island." An advantage of the argument presented here is that it provides a quantitative way to see global symmetry violation just *after the Page time*, and does not rely on the physics close to the final stage of the black hole evaporation.

Notice that if we gather the entire radiation of a completely evaporated black hole, it will be in a pure state. The pure state will be a superposition of states with different charge. Hence we expect that the relative entropy after the black hole has completely evaporated to formally diverge. On the other hand, the Rényi versions of the relative entropy that we study in section 3 will still be well defined and give sensible answers. A related point is discussed in appendix C, where we explore global symmetry violations in a different model of JT gravity.

Instead of starting with a state that is annihilated by the charge, one could also imagine forming the black hole from a general initial state ρ_0 that is not invariant under the global symmetry U that acts on the entire system.[8] In this case, one could still see that global symmetry is violated as follows. If the symmetry were exact, then when we look at any

---

[7]In the 1-sided zero-temperature case, the scrambling time is infinite but the $t_{\text{island}} \sim \phi_r/c$.

[8]Here we are assuming that before the black hole forms, all the matter are far apart and the gravity effect is weak enough such that the operator U is well defined.

subsystem R at later time t, we should always find

$$
\begin{aligned}
S\left(U_R\rho_R(t)U_R^\dagger\Big|\rho_R(t)\right) &\leq S\left(Ue^{iHt}\rho_0 e^{-iHt}U^\dagger\Big|e^{iHt}\rho_0 e^{-iHt}\right) \\
&= S(U\rho_0 U^\dagger|\rho_0),
\end{aligned}
\tag{2.29}
$$

where on the first line we are using the monotonicity of relative entropy under restricting to a subsystem R, and we used the assumption that U is an exact symmetry to get to the second line. However, if we use the gravitational formula to compute $S\left(U_R\rho_R(t)U_R^\dagger\Big|\rho_R(t)\right)$, we expect to get an increasing function with t after the Page time,[9] so it cannot be bounded by a constant $S(U\rho_0 U^\dagger|\rho_0)$ as in (2.29), which necessarily means that the global symmetry is violated.

3 Global symmetry violation and replica wormholes

3.1 Charge flowing through the replica wormhole

As we've seen in section 2, when there is an island in the entanglement wedge of R, the global symmetry violation is reflected in the relative entropy. In the computation of the von Neumann entropy using the replica trick, the island arises from the replica wormholes [19, 20]. So it is natural to consider the replica version of the relative entropy, and ask what the role of replica wormholes are in these quantities. Similar to the von Neumann entropy, the relative entropy can be computed via a replica trick [53] by

$$
S(\rho|\sigma) = \lim_{n\to 1}\frac{1}{1-n}\mathrm{tr}\left[\rho(\sigma^{n-1}-\rho^{n-1})\right].
\tag{3.1}
$$

To be concrete, we will take the two-sided black hole setup in [28] as an example, whose replica wormhole solution has been constructed explicitly in certain limits in [19]. The goal of this section is not to study the replica wormhole geometries in detail, but to understand what features of the replica wormholes are related to charge violation. We will be considering the thermofield double state $|\mathrm{TFD}\rangle$ of the system, and study the density matrix of a region R which contains part of both the left bath and the right bath (see figure 4). Depending on the parameters, we can have an island I that is slightly outside the horizon already at $t=0$; our arguments in section 2 would then apply.

Here we would like to consider the $n=2$ replica:

$$
\frac{Z_2(\alpha)}{Z_1^2} = \mathrm{tr}\left[U_R\rho_{\mathrm{exact}}(R)U_R^\dagger\rho_{\mathrm{exact}}(R)\right], \quad U_R = e^{i\alpha Q_R},
\tag{3.2}
$$

where we take the symmetry to be an U(1) global symmetry just for simplicity. We also divided by Z_1^2 where Z_1 is given by the path integral on a single replica, to get the expression for a normalized density matrix. If the symmetry were exact, then we should

---

[9]We expect the curve to be qualitatively similar to figure 2, if we replace the size of the interval by time t. The reason is that there will be more and more correlated charge fluctuations between I and the Hawking radiation R (this is similar in spirit to figure 3). It would be interesting to understand the calculation for evaporating black holes in more detail.
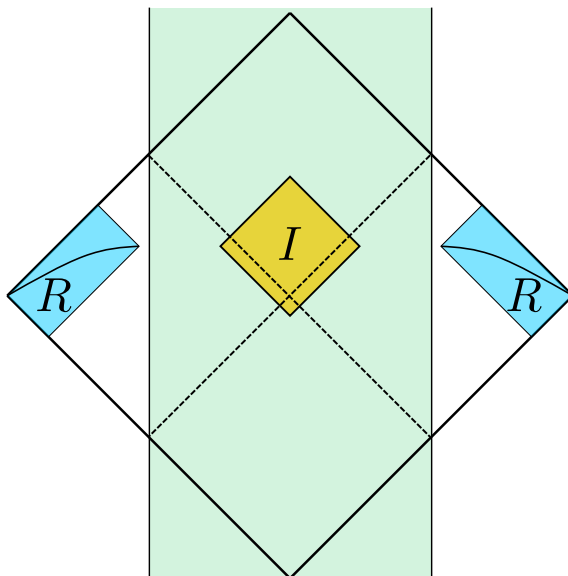
Figure 4. A two-sided black hole in JT gravity coupled to two bath regions [19]. When we compute the entropy of a large region R in the bath, depending on the parameters, we can have an island I in the gravity region. At sufficiently late times $t > 0$ there will always be an island.
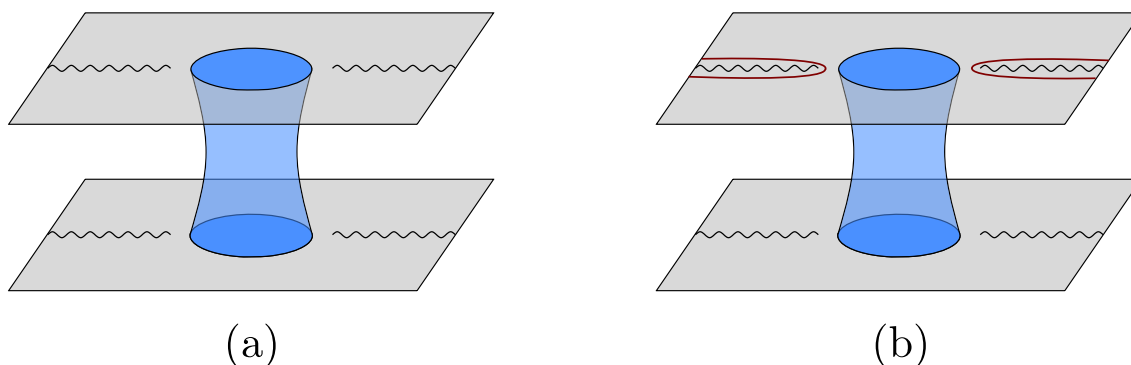

Figure 5. (a) The replica wormhole geometry $\mathcal{M}_2$ for computing $Z_2(0)$. (b) For $Z_2(\alpha)$, we have the symmetry transformation operators inserted around the cuts in one of the replicas.

have $Z_2(\alpha) = Z_2(0)$, for the reasons discussed in section 2.1. As explained in [19], $Z_2(0)$ can be computed via a gravitational path integral with two replicas, and in the case with an island, the path integral is dominated by a replica wormhole geometry (see figure 5(a)), which we denote by $\mathcal{M}_2$. The only difference in the computation of $Z_2(\alpha)$ is that we have the extra insertions of U_R and $U_R^\dagger$, which are represented by the red lines in figure 5(b).

Importantly, since we are computing (3.2) in the semiclassical description that has the global symmetry, the red lines in figure 5(b) represent topological operators.[10] This means we can deform the curves arbitrarily as long as they do not cross any charged operators (in which case they would pick up a phase). Since we are considering the thermofield double

---

[10]The topological operator is well-defined for other symmetries as well, including non-Abelian cases and discrete symmetries.
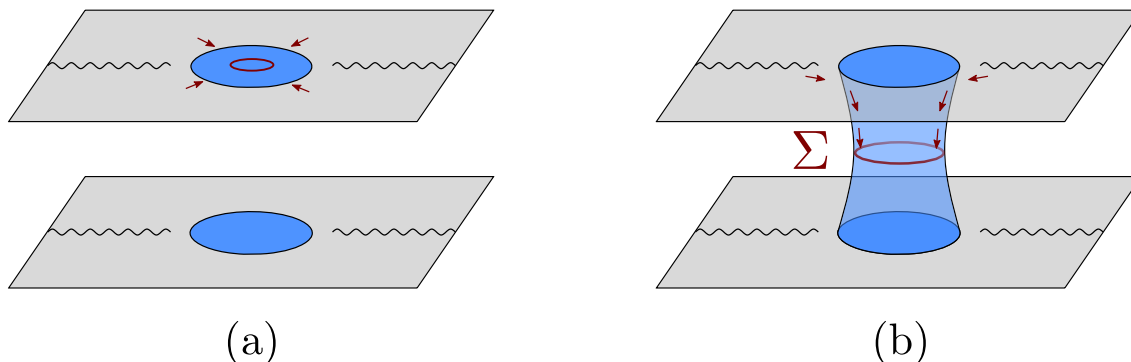
Figure 6. (a) Without the replica wormhole, the topological operator can be shrunk to a point. (b) With the replica wormhole, we can deform the topological operator to surround the throat of the wormhole. Therefore the dependence on the parameter α diagnoses the charge flowing through the wormhole.

state, which does not contain any insertions, we can deform the topological operator freely. If there were no replica wormholes (see figure 6(a)), we could then deform and shrink the topological operator to a point, which tells us

$$\frac{Z_2(\alpha)}{Z_2(0)} = 1, \quad \text{(no replica wormhole).} \tag{3.3}$$

On the contrary, with the replica wormhole, the topological operator cannot be shrunk to a point, but it can be deformed to the throat Σ of the wormhole, see figure 6(b). So we have

$$\frac{Z_2(\alpha)}{Z_2(0)} = \langle e^{i\alpha Q_{\text{wormhole}}} \rangle_{\mathcal{M}_2}, \quad \text{(with replica wormhole),} \tag{3.4}$$

where $Q_{\text{wormhole}} = \int_\Sigma *J$ and J^μ is the Noether current. This is the finite replica analogue of equation (2.12). In general, we expect that $\langle e^{i\alpha Q_{\text{wormhole}}} \rangle_{\mathcal{M}_2} < 1$, simply because the value of Q_{wormhole} can fluctuate in the path integral. Thus we would have

$$Z_2(\alpha) < Z_2(0), \tag{3.5}$$

which means that the density matrix $\rho_{\text{exact}}(R)$ is not invariant under the symmetry transformation, as we concluded by studying the relative entropy.

We can use the intuition from a weakly-coupled theory to understand this effect slightly better. If we consider a small (meaning contractible) loop of virtual particles near the throat of the geometry, the net charge flow will be zero, since on a spatial slice Σ, we will have one particle and one anti-particle, see figure 7(a). However, the virtual loop could be non-contractible due to the cuts in the bath region, see figure 7(b).

3.2 Off-diagonal contributions to the density matrix

The fact that $U_R \rho_{\text{exact}}(R) U_R^\dagger$ and $\rho_{\text{exact}}(R)$ are different indicates that $\rho_{\text{exact}}(R)$ must have some off-diagonal elements in the charge basis. We can try to "open up" the quantity
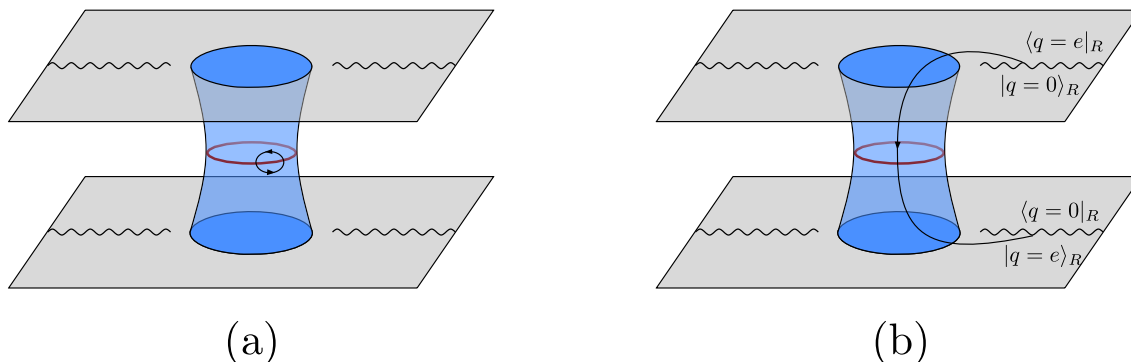
Figure 7. Two virtual processes that contribute to $Z_2(\alpha)$. (a) The smaller, contractible loop does not give a contribution that depends on α, since on the red circle there is both a particle and an anti-particle. (b) The larger loop is non-contractible, contributes to the off-diagonal part $\mathrm{tr}\,[\rho_{e,0}\rho_{e,0}]$ (see section 3.2), and therefore to the α dependence. This should be compared with figure 3.

in (3.2) and look at the contribution to the off-diagonal elements more explicitly:

$$\mathrm{tr}\left[e^{i\alpha Q_R}\rho_{\mathrm{exact}}(R)e^{-i\alpha Q_R}\rho_{\mathrm{exact}}(R)\right] = \sum_q \mathrm{tr}\left[\rho_{q,q}^2\right] + e^{i\alpha(q-q')}\sum_{q\neq q'}\mathrm{tr}\left[\rho_{q,q'}\rho_{q',q}\right], \qquad (3.6)$$

where

$$\rho_{q,q'} \equiv P_q\rho_{\mathrm{exact}}(R)P_{q'} \qquad (3.7)$$

and $P_q \equiv \sum_i |q,i\rangle\langle q,i|$ is the projector into the subspace with $Q_R = q$, with i running over the states in the same-charge subspace. Note that the projection only acts on the fields that transform under the symmetry, which we take to be a small subset of the fields. Since we are tracing out all the other fields, we expect a similar replica wormhole solution for each term in (3.6) independently. We see from (3.6) that all the dependence on α comes from the off-diagonal terms in the density matrix. In this section, we give a more direct explanation of the appearance of off-diagonal terms, and how they arise from Euclidean wormholes in gravity.

One way to understand the off-diagonal term $\mathrm{tr}\,[\rho_{q,q'}\rho_{q',q}]$ is as follows. We can rewrite the quantity as
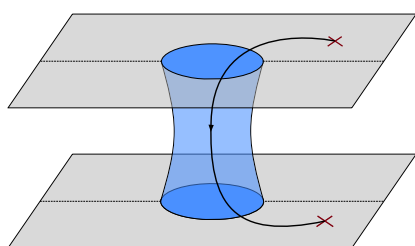
$$\mathrm{tr}\,[\rho_{q,q'}\rho_{q',q}] = \sum_{i,j} |\langle\mathrm{TFD}|q,i\rangle\langle q',j|\mathrm{TFD}\rangle|^2, \qquad (3.8)$$

and in the path integral, it corresponds to putting boundary conditions for the fields along the branch cuts such that they have definite charges. (This should not be taken too literally, as such a state might be a very high energy.) In the semiclassical approximation, we can think of the boundary conditions as places where the worldline of a charged particle can be created or destroyed. Let's take $\mathrm{tr}\,[\rho_{e,0}\rho_{0,e}]$ as a simple example, where e is the unit charge of a charged particle. Due to the boundary condition, we will have a worldline starting from the bra part of the cut in the first replica, and the only place where it can land on is the ket part of the cut in the second replica. The only possibility for the contribution to be nonzero is that the charge can go through the replica wormhole, as shown in figure 7(b).

Despite that we have $\mathrm{tr}\left[\rho_{q,q'}\rho_{q',q}\right] \neq 0$, if we compute $\mathrm{tr}\left[\rho_{q,q'}\right]$ with $q \neq q'$ using gravity, we will find a zero answer, up to nonperturbative corrections. This is because we only have one replica, and there is nowhere the charged particle can go to. This effect can be viewed as a special case of a more general phenomena involving Euclidean wormholes. Consider a state of the effective quantum field theory in the bath obtained by inserting some charged operators in the thermofield double state $|\psi_q\rangle = O_{q_1}O_{q_2}\cdots O_{q_n}|\mathrm{TFD}\rangle$. Here $O_q = O_q(\tau, x)$ is some local operator with charge q evaluated at a general Euclidean or Lorentzian time. Ignoring the effects of higher topologies, this state will be orthogonal to the state $|\psi_{\tilde{q}}\rangle = O_{\tilde{q}_1}O_{\tilde{q}_2}\cdots O_{\tilde{q}_m}|\mathrm{TFD}\rangle$ unless the total charges are equal $q = \sum_i q_i = \tilde{q} = \sum_i \tilde{q}_i$, namely $\langle\psi_q|\psi_{\tilde{q}}\rangle = 0$ up to exponentially small errors. However, wormhole geometries will lead to a small squared-overlaps when $q \neq \tilde{q}$ [54]:

$$|\langle\psi_q|\psi_{\tilde{q}}\rangle|^2 = \delta_{q-\tilde{q}}\,(\text{disconnected}) + e^{-S_0}(\text{connected}) \tag{3.9}$$

The connected contribution comes from an Euclidean wormhole, where the charged particles go through the wormhole. For example, the square of a 1-pt function $\langle O_q\rangle$ with charge q is given by

$$|\langle\mathrm{TFD}|O_q|\mathrm{TFD}\rangle|^2 \quad = \quad \tag{3.10}$$

Note however that (3.10) might require an ensemble average interpretation, since the left hand side is the square of $\langle\mathrm{TFD}|O_q|\mathrm{TFD}\rangle$, while the right hand side does not factorize manifestly. In a single unitary system, the proper interpretation of (3.10) might be that the wormhole computes the average of $|\langle\mathrm{TFD}|O_q|\mathrm{TFD}\rangle|^2$ among a suitable family of operators with charge q. We refer the readers to [54–64] for more discussions on these issues about Euclidean wormholes. We should stress that the quantity $\mathrm{tr}\left[\rho_{q,q'}\rho_{q',q}\right]$ does not factorize in the first place, since each charge subspace still contains a huge number of states that we sum over. So the quantity we compute does not suffer from factorization problem immediately. Due to the summation over a large number of intermediate states, we have a replica wormhole geometry which is a saddle point solution that dominates the path integral, while in general the wormhole in (3.10) will be an off-shell configuration (see however an on-shell construction in [54]).

If we imagine that the wormhole region is far away from the operator insertions, we can summarize the effects of the wormhole using an effective picture. If we think of the gravity region as a complicated boundary condition for the bath fields, this boundary condition will break explicitly the global symmetry. If the gravity theory has a holographic description as we've assumed in section 2, then this boundary condition would be realized by the coupling between the quantum mechanical system and the bath CFT. More generally, we expect the same picture to hold even when the gravity theory does not have a holographic description.
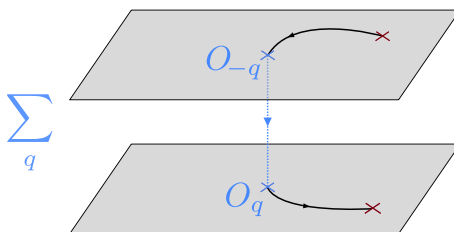
Figure 8. We can reproduce the effects of the wormhole by an effective picture where we sum over correlated operators. This is also what happens from the boundary point of view, where the gravity region is replaced by a boundary condition which completely breaks the global symmetry. Here we are imagining that the operator insertions in the bath are far away from the gravity region. In some theories with a disorder average, the dashed blue line may represent a contraction of couplings.

Far away from this boundary condition, the circle is hard to distinguish from a point, so we can replace the boundary condition by a sum of operator insertions, see figure 8. We can imagine that there is a small coefficient for some operator $c_q O_q$ where c_q has mean zero in some appropriate averaged sense. Then ignoring these operators is equivalent to only considering disconnected geometries. When we include connected geometries, we are taking into account the non-zero variance of these coefficients $|c_q|^2 \sim e^{-S_0}$. Of course, this is just the old story about "α-parameters" and Euclidean wormholes [1, 61, 65, 66], what we did in this section is making the relation to replica wormholes more explicit.

From this perspective, it is clear that a Euclidean wormhole with charge $\pm q$ propagating through the throat is giving rise to a state with charge $\pm q$ different than the "naive" charge of the state (e.g. the charge of the state we would have assigned in the effective description of the quantum fields on the trivial topology).

Now let us return to the density matrix of the fields in a subregion R. We will slightly generalize the previous discussion by considering a state $|\psi_q\rangle$ which is obtained by inserting some charged operators in the Euclidean evolution. Let us denote elements of a charge eigenbasis by a composite index r such that $|r\rangle = |q_r, i\rangle$, where i runs over states with the same charge q_r. Then,

$$
\begin{aligned}
\rho_{\text{exact}}(R) &= \operatorname{tr}_{\bar{R}} |\psi_q\rangle \langle \psi_q| , \\
Z_2(\alpha)/Z_1^2 &= \sum_{r,s} \langle r| e^{-i\alpha Q} \rho_{\text{exact}}(R) e^{i\alpha Q} |s\rangle \langle s| \rho_{\text{exact}}(R) |r\rangle \\
&= \sum_{r,s} e^{i\alpha(q_s - q_r)} \langle \psi_q| O_{s,r}^R |\psi_q\rangle \langle \psi_q| O_{r,s}^R |\psi_q\rangle \\
&= \sum_{r,s} e^{i\alpha(q_s - q_r)} |\langle \psi_q|\psi_{\tilde{q}}\rangle|^2 \\
|\psi_{\tilde{q}}\rangle &= O_{r,s}^R |\psi_q\rangle , \quad \tilde{q} = q + q_r - q_s .
\end{aligned}
\tag{3.11}
$$

where $O_{r,s}^R \equiv |r\rangle \langle s|$ is an operator with charge $q_r - q_s$. Neglecting gravity, we would have expected to only receive a contribution when $q_r = q_s$. However, from (3.9) we see that each off-diagonal contribution to the wormhole corresponds to diagrams where particles carrying total charge $q_r - q_s$ propagate through the wormhole, see figure 9.
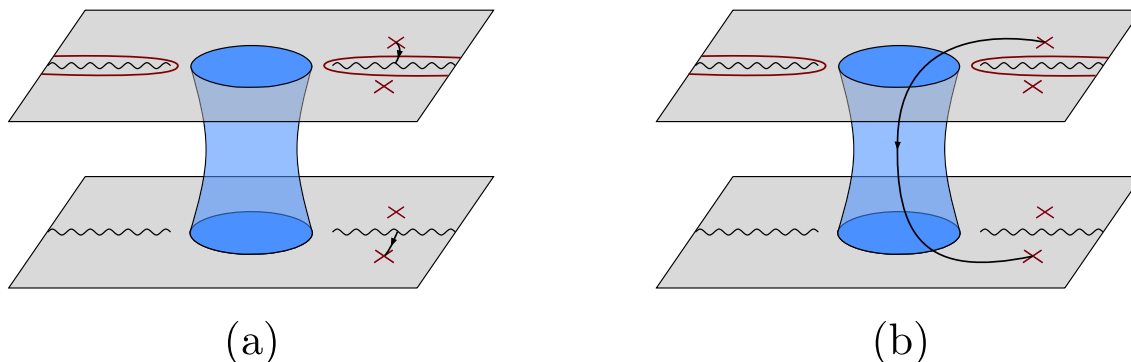
Figure 9. We consider a global state which is not the vacuum but contains some operator insertions in the bath region. (a) The diagonal contribution to the density matrix comes from particles propagating through the cuts. (b) Off-diagonal contributions come from world-lines with non-zero net charge propagating through the wormhole.

As a side remark, let us comment on the relation between scattering in the black hole background. We have seen that the same effect which gives the non-zero relative entropy is also responsible for non-zero values of correlation functions which would naively vanish from symmetry considerations. These correlation functions in the bath are closely related to scattering in the black hole background. For example, a 4-pt function in the bath region in the appropriate out-of-time-order configuration could be interpreted as the traversable wormhole signal [67] that violates global symmetry.

3.3 Comments on gauge symmetries

While exact global symmetries are not allowed in quantum gravity, gauge symmetries of course can exist. Here we explain the difference between the two from the replica wormhole point of view. To be concrete, we will consider the same set up as in (3.1), where the CFT in the flat space region has an U(1) global symmetry, while the difference is that the symmetry is gauged in the gravity region. By the same consideration as in (3.1), we would still have

$$\frac{Z_2(\alpha)}{Z_2(0)} = \langle e^{i\alpha Q_{\text{wormhole}}} \rangle_{\mathcal{M}_2}, \quad \text{(with replica wormhole)}. \tag{3.12}$$

However, in the path integral, there will be an integral over the spatial zero mode of the gauge field $\int dA_0 \exp\left(iA_0 \int_\Sigma j^0\right) = \delta(Q_{\text{wormhole}})$ which sets the charge propagating through the wormhole to zero. Thus we have

$$Z_2(\alpha) = Z_2(0) \tag{3.13}$$

for arbitrary α as enforced by the gauge constraint. Similarly we have $Z_n(\alpha) = Z_n(0)$ for higher replicas as well. It then follows that as $n \to 1$ the relative entropy must vanish. This shows that the symmetry is exact in the boundary description as expected.[11]

---

[11]It would be interesting to understand how to see the vanishing of the relative entropy directly from the relative entropy formula (2.9), without going through the replica argument.

We can rephrase this argument slightly differently as follows. Consider $(d+1)$-dimensional electromagnetism on a compact space with no boundary $\Sigma \times$ time. In such a theory, it is clear that the overall charge on Σ must vanish. A net charge would source electric field lines, which have nowhere to go. Integrating Gauss's law $\nabla \cdot E = \rho$ on a spatial manifold Σ with no boundary, we get the constraint that $Q = 0$. If we think of the direction along the Euclidean wormhole as time, this makes it clear that no net charge can propagate through the wormhole. This is true even when the gauge coupling goes to zero. Note that if we view gravity as a gauge theory, the analogous equation says that the ADM Hamiltonian of a closed universe vanishes. If we perform a Kaluza-Klein reduction of some of the compact dimensions along the wormhole, the momenta in the compact dimensions will become gauge charges. So we will get a constraint that the total momentum in each compact dimension vanishes.

4 Discussion and conclusion

It has been suggested for a long time that Euclidean wormholes lead to symmetry violation since charged particles may propagate through the wormhole. Normally, this is an exponentially small effect. The main point of this paper is showing that when we compute certain relative entropies, the tiny effects add up and give an $\mathcal{O}(1)$ result. This is morally similar to what happens in the Page curve after the Page time, where the tiny correlations in the Hawking radiation bring down the entropy. The novelty here is that instead of computing the entropy, we are computing a relative entropy which involves symmetry generators acting in the bath region, so these quantities tell us about the symmetry violation. Importantly, this effect can be computed with just knowledge of the semi-classical description. Our results are based on the recent paradigm of islands and replica wormholes. The same logic which demands that replica wormholes be included in the computation of the Page curve also demands that they should be included in computations of the relative entropy. Of course, there are unresolved puzzles about the inclusion of Euclidean wormholes in gravitational path integrals [54–61, 63, 64]. In some cases, it is known that Euclidean wormholes are computing a disorder average over theories. In these cases, it is possible that the ensemble may have a symmetry group G that each individual theory does not. (This means that the probability distribution over theories is invariant under the action of G on theory-space). It is tempting to say that when there is an ensemble interpretation, if we have a matter theory with global symmetry group G coupled to gravity, the group G should be re-interpreted as the symmetry of the ensemble or the third-quantized theory. See appendix C.3 for some extra comments.

The relative entropy $S(U_R \rho_{\text{exact}}(R) U_R^\dagger | \rho_{\text{exact}}(R))$ that we considered involves only the density matrix in the bath region. Although it is not a direct observable, it might be able to provide contraints or bounds on physical observables, which requires further investigation. If one can create (or simulate holographically) black holes in the lab, then the relative entropy and its Rényi versions are in principle measurable for observers who have multiple copies of the system. In principle, measuring the say $n = 2$ version of the relative entropy would not be hard (in the sense of computational complexity) if one had access to $n = 2$

copies of the system. We do not need to measure each component of the density matrix independently.

We stress that it is surprising that a quantity that characterizes the nonperturbative violation of global symmetry can be computed in the semiclassical description. This suggests that in some sense, the way gravity violates global symmetry is far from generic. By this we mean the following: if we take some quantum field theory and put a random boundary condition that contains tiny terms that breaks any global symmetries, we would not expect the exact relative entropy to satisfy any formula given by the effective theory where we have erased the tiny terms. Of course, the surprise we are referring to is similar to the surprise that semiclassical gravity knows about the Page curve.

We've focused on idealized situations where there is no global symmetry violation within the semiclassical description. However, the argument we had can be easily generalized to situations where the symmetry has already been violated semiclassically. (For example, there could be Planck-suppressed operators in the Lagrangian which break explicitly any global symmetries.) In such theories, the relative entropy of the semiclassical density matrix will already be non-zero, i.e. $S\left(U_R\rho_{\text{semi}}(R)U_R^\dagger\Big|\rho_{\text{semi}}(R)\right) > 0$. However, if we consider the relative entropy of the exact density matrix, as long as the island still exists, we will have

$$\begin{aligned}
S\left(U_R\rho_{\text{exact}}(R)U_R^\dagger\Big|\rho_{\text{exact}}(R)\right) &= S\left(U_R\rho_{\text{semi}}(R \cup I)U_R^\dagger\Big|\rho_{\text{semi}}(R \cup I)\right) \\
&> S\left(U_R\rho_{\text{semi}}(R)U_R^\dagger\Big|\rho_{\text{semi}}(R)\right),
\end{aligned} \tag{4.1}$$

where we used the monotonicity of relative entropy from the first line to the second line. The difference between the exact answer and the semiclassical answer will still be order one due to the inclusion of the island. Thus for an evaporating black hole, we will still find a sudden increase of the relative entropy at the Page time, which is a clear signature of the global symmetry violation from nonperturbative effects.

We have emphasized that the symmetry violating effects we are considering are related to Euclidean wormholes. On the other hand, it has long been known that the Hawking process violates symmetries; these previous arguments only involved the naive disconnected geometries. In order to emphasize what our new arguments buy, let us return to the setup discussed in section 2.3. The matter theory in the bulk is a large c CFT + a probe massive field with a U(1) symmetry and particles of mass m. Let us consider a large black hole which has a Hawking temperature $T \ll m$ that is older than the Page time. Imagine that it formed from neutral matter. Now consider the same experiment in section 2.3. If a few charged particles fall into the black hole, we might wonder whether the experimentalist which remains outside the hole can see charge violation. Naively to check the argument of [4], she would have to wait an extremely long time for the black hole to evaporate. In particular, she would have to wait for the black hole to become small enough that $T \gg m$, at which point it could start to radiate the U(1) charge.

However, our proposal is that even after a scrambling time, the clever experimentalist can in fact report a violation of charge conservation by measuring off-diagonal components

of the density matrix, or if she can control multiple copies of the system, a suitable Rényi relative entropy.

Note that our result is similar in spirit to the result of Hayden and Preskill [22]. Naively, one would have guessed that to recover the information from a diary thrown into the hole, one would have to wait until almost the end of the evaporation process. But [22] showed that one only needs to wait for a scrambling time if the black hole is older than the Page time. Here we are in some sense refining their result and showing that the information in the Hawking radiation recovered after the scrambling time is sufficient to demonstrate that charge conservation is violated.[12]

The connection between the appearance of islands and violation of global symmetries implies that we can see clear signatures of symmetry violation even in situations where charge conservation considerations like those in [4] do not immediately apply. Besides the Hayden-Preskill-like experiment discussed above, an example is the extremal (zero temperature) black hole. Another case is the finite temperature black hole in equilibrium with a bath that never evaporates. We expect our results to generalize to the cosmological setups of [35, 36, 40] where the old arguments would not establish a clear signature of symmetry violation. Of course, the old arguments only involve measuring simple observables like the total charge of the Hawking radiation. In our work, we are discussing the relative entropy which is a much more complicated quantity.

It would be interesting to see if our arguments can be generalized to address questions about gauge charges like the weak gravity conjecture, completeness conjecture, etc. [3, 69].

Acknowledgments

A Nonperturbative correlation between subsets of fields in the Hawking radiation

As we mentioned in section 2, the island formula implies that there exists substantial correlation between different fields in the Hawking radiation. In this appendix, we illustrate this point in more details. Instead of dividing the Hawking radiation into spatial subsystems, we can also divide the Hawking radiation in a fixed region into subsets of fields. In other words, we are dividing the Hilbert space of some subregion into the tensor product of some smaller Hilbert spaces. This division is at least well defined when the Hawking radiation is in an asymptotic region, where we can neglect the dynamical gravity effects.

---

[12]Despite the analogy, we should not confuse the ability to recover the charge of the particle using entanglement reconstruction with the effect that we are talking about here. The violation of charge conservation is more directly related to the fact that naively orthogonal black hole states are not orthogonal in the exact description, this is particularly clear in the model of appendix C. See also [68] for another recent discussion on the interplay between global symmetry and the Hayden-Preskill protocol.

Let's consider a region R that is big enough such that we have islands. Now we separate the fields in R into a large portion (labeled by L) and a small portion (labeled by S). We take the number of degrees of freedom in S to be much smaller than L. In the example of section 2.2, S will be the massless fermion field under the U(1) global symmetry, while L will be the rest of the fields.

We could compute the entropies of the three density matrices $\rho_{\text{exact}}(R_L \cup R_S)$, $\rho_{\text{exact}}(R_L)$ and $\rho_{\text{exact}}(R_S)$ using the QES prescription. For $\rho_{\text{exact}}(R_L \cup R_S)$ and $\rho_{\text{exact}}(R_L)$, we will find an island in the entanglement wedge:

$$S(\rho_{\text{exact}}(R_L \cup R_S)) = \frac{A(\partial I)}{4G_N} + S(\rho_{\text{semi}}(R_L \cup R_S \cup I_L \cup I_S)), \tag{A.1}$$

$$S(\rho_{\text{exact}}(R_L)) = \frac{A(\partial I)}{4G_N} + S(\rho_{\text{semi}}(R_L \cup I_L \cup I_S)), \tag{A.2}$$

where we neglected the change in the position of the island due to the addition of the field S, since it will only give subleading corrections. Importantly, in (A.2), one gets contributions from both I_L and I_S in the island. This is because the island should be for all the fields, as it follows from the replica wormhole geometry which all the fields live on. (In other words, the emergent twist operators act on all fields.) For $\rho_{\text{exact}}(R_S)$, since its semi-classical entropy is small, there will not be an island:

$$S(\rho_{\text{exact}}(R_S)) = S(\rho_{\text{semi}}(R_S)). \tag{A.3}$$

Combining (A.1) to (A.3), one finds

$$I_{\text{exact}}(R_S : R_L) = I_{\text{semi}}(R_S : R_L \cup I_L \cup I_S). \tag{A.4}$$

In particular, by the monotonicity of mutual information, one has

$$I_{\text{exact}}(R_S : R_L) > I_{\text{semi}}(R_S : R_L), \tag{A.5}$$

and the difference will be an order one amount. The difference comes from nonperturbative effects that are not present in the semiclassical description. Notice that in the vacuum state, the semiclassical mutual information $I_{\text{semi}}(R_S : R_L)$ will vanish in the limit that S and L are decoupled in the semiclassical theory. On the other hand, the right hand side of (A.4) reduces to $I_{\text{semi}}(R_S : I_S)$, which is nonzero and finite even when the coupling between S and L is negligible in the semiclassical description.

B Details on the relative entropy for massless free fermion in 2D

As derived in [49, 70], for a two dimensional free massless fermion in the vacuum state, the modular Hamiltonian of the union of $I = (a_1, b_1)$ and $R = (a_2, b_2)$ takes the following form:

$$K = K_{\text{loc}} + K_{\text{noloc}},$$

$$K_{\text{loc}} = 2\pi \int_{I \cup R} dx \, \omega'(x)^{-1} T(x),$$

$$K_{\text{noloc}} = i2\pi \int_{I \cup R} dx \, \psi^\dagger(x) \frac{(b_1 - a_1)(a_2 - b_1)(b_2 - a_1)(b_2 - a_2)}{\omega'(x)^2 (x - a_1)(x - a_2)(x - b_1)(x - b_2)} \tag{B.1}$$

$$\frac{1}{x(a_1 + a_2 - b_1 - b_2) + (b_1 b_2 - a_1 a_2)} \psi(\bar{x}),$$

where $T(x) = \frac{1}{2}\left[i\partial_x\psi^\dagger(x)\psi(x) - \psi^\dagger(x)i\partial_x\psi(x)\right]$ is the energy density operator, and $\omega(x), \bar{x}$ are given by

$$\omega'(x) = \frac{1}{x - a_1} + \frac{1}{x - a_2} - \frac{1}{x - b_1} - \frac{1}{x - b_2}$$
$$\bar{x} = \frac{a_1 a_2\left(x - b_1 - b_2\right) - b_1 b_2\left(x - a_1 - a_2\right)}{x\left(a_1 + a_2 - b_1 - b_2\right) + \left(b_1 b_2 - a_1 a_2\right)}. \tag{B.2}$$

The formulas are the same for ψ_+ and ψ_-, so we've omitted the chirality of the fermion. As discussed in (2.20), we are interested in computing the expectation value $\langle 0| K_{\text{noloc}} |0\rangle$. Using

$$\langle 0| \psi^\dagger(x)\psi(y) |0\rangle = \frac{i}{2\pi}\frac{1}{x - y}, \tag{B.3}$$

when $x \neq y$, we can compute the integral in (B.1), and find

$$\langle 0| K_{\text{noloc}} |0\rangle = -\frac{1}{4}\left(1 + \frac{(2\eta - 1)\arctan\sqrt{\frac{\eta}{1-\eta}}}{\sqrt{\eta(1-\eta)}}\right), \tag{B.4}$$

where the cross ratio η is defined as

$$\eta \equiv \frac{(b_1 - a_1)(b_2 - a_2)}{(a_2 - a_1)(b_2 - b_1)}. \tag{B.5}$$

When computing the integral, it is convenient to set $\{a_1, b_1, a_2, b_2\} = \{0, \eta, 1, \infty\}$. (B.4) is the contribution from one of the chiral modes, so to get the total contribution, we multiply it by two.

C Global symmetry in JT gravity + EoW branes

In this section, we explore issues related to global symmetry violations in the "West coast" model [20], namely JT gravity with end-of-the-world (EoW) branes. We refer the reader to [20] for definitions and conventions. Some similar comments could be made in the simplified model of [61].

C.1 Coarse-grained entropy

A natural question is whether one can define a coarse-grained density matrix $\tilde{\rho}_R$ of the radiation, whose entropy does not follow the Page curve but instead is given by the naive, non-minimal quantum extremal surface. In the setup of [20], we want the coarse-grained entropy to be $\approx \log k$, even when k is bigger than e^{S_0}. Here we point out that if we have a large symmetry group, we can use group-averaging to define a coarse-grained density matrix:

$$\tilde{\rho}_R = \int dg\, \mathrm{U}(g)\rho_R\mathrm{U}(g^{-1}). \tag{C.1}$$

Note that the group-averaging is a completely positive trace preserving (CPTP) map. It takes any density matrix into a density matrix which is invariant under the symmetry group. The CPTP property ensures that

$$S(\tilde{\rho}_R) \geq S(\rho_R). \tag{C.2}$$

For a simple model where we can explicitly compute $S(\tilde{\rho}_R)$, consider the west coast model [20]. From the bulk (semi-classical) perspective, the model is defined by JT gravity + end-of-the-world (EoW) branes. The EoW branes carry an index i such that

$$\langle i|j\rangle_{\text{bulk}} = \delta_{ij}.\qquad\qquad(\text{C.3})$$

This equation should not be confused with a statement about the actual boundary states that correspond to these branes, which are not exactly orthonormal. The content of this equation is simply to prescribe rules for evaluating the bulk path integral. The point we would like to make here is that the bulk theory at the semi-classical level has a U(k) global symmetry, where the EoW branes transform in the fundamental.

Now let us consider the replica wormhole computation of $S(\tilde{\rho}_R)$. We will start by computing the Rényi entropies. Let us consider the 2-replica computation of $\text{tr}\,(\tilde{\rho}_R^2)$. Since the overall state of the black hole and the radiation is an entangled state

$$|\Psi\rangle = \frac{1}{\sqrt{k}} \sum_{i=1}^{k} |\psi_i\rangle_{\text{B}} |i\rangle_{\text{R}},\qquad\qquad(\text{C.4})$$

acting with the symmetry generator on the radiation $|i\rangle \to U |i\rangle_{\text{R}} = U_{ji} |j\rangle_{\text{R}}$ is equivalent to acting on the black hole states $|\psi_i\rangle_{\text{B}} \to \sum_j U_{ji} |\psi_j\rangle_{\text{B}}$. Clearly if we consider the disconnected saddle, the symmetry generators act trivially. But for the connected saddle, we actually get the constraint $i = j$. So whereas

$$
\begin{aligned}
\text{Tr}\left(\rho_{\text{R}}^2\right) &= \frac{kZ_1^2 + k^2 Z_2}{(kZ_1)^2} = \frac{1}{k} + \frac{Z_2}{Z_1^2}, \\
\text{Tr}\left(\tilde{\rho}_{\text{R}}^2\right) &= \frac{kZ_1^2 + kZ_2}{(kZ_1)^2} = \frac{1}{k}\left(1 + \frac{Z_2}{Z_1^2}\right).
\end{aligned}
\qquad\qquad(\text{C.5})
$$

Notice that the connected saddle never dominates, no matter how big k is.

In fact, from the boundary point of view we can compute the exact density matrix after group averaging:

$$
\begin{aligned}
\rho_{\text{R}} &= \frac{1}{k} \sum_{i,j=1}^{k} |j\rangle \langle i|_{\text{R}} \langle\psi_i|\psi_j\rangle_{\text{B}}. \\
\tilde{\rho}_{\text{R}} &= \frac{1}{k} \sum_{i,j=1}^{k} \int dU\, U |j\rangle \langle i|_{\text{R}} U^\dagger \langle\psi_i|\psi_j\rangle_{\text{B}} \\
&= \frac{1}{k} \sum_{i} |i\rangle \langle i|_R \langle\psi_i|\psi_i\rangle \\
&\approx \frac{1}{k} \sum_{i} |i\rangle \langle i|_R.
\end{aligned}
\qquad\qquad(\text{C.6})
$$

Note that this is almost a completely mixed density matrix. (It is not perfectly mixed because the norms of the states $\langle\psi_i|\psi_i\rangle$ have small fluctuations. However, to leading order in e^{-S_0}, this effect is negligible.) Notice that at large k, this gives a concrete meaning to a sub-dominant QES.

C.2 Relative entropy

Let us start with computing the relative purity $\operatorname{tr} U \rho U^\dagger \rho$ in the West coast model. Up to a normalization,

$$
\begin{aligned}
\operatorname{tr} U \rho U^\dagger \rho = U_{ij} \rho_{jk} U_{kl}^\dagger \rho_{li} = \\
= U_{ij} U_{kl}^\dagger \langle \psi_j | \psi_k \rangle \langle \psi_l | \psi_i \rangle \\
= \sum_{i,k} \delta_{ik} Z_1^2 + U_{ii} U_{kk}^\dagger Z_2
\end{aligned}
\tag{C.7}
$$

Notice that the unitary does not affect the disconnected contribution. Furthermore, when we set the group element to the identity $U = I$, we recover the result in [20].

Now we can also use the QES prescription to compute the relative entropy. At small k, there is no island and the QES prescription gives 0. Since the semi-classical density matrix is a pure state:

$$
\rho_{\text{semi}} = \sum_{i,j} (|\psi_i\rangle \langle \psi_j| \otimes |i\rangle \langle j|)
\tag{C.8}
$$

the relative entropy would be infinite if $U \neq 1$:

$$
S(U_R \rho_{\text{semi}}(R \cup I) U_R^\dagger | \rho_{\text{semi}}(R \cup I)) = \infty, \quad U \neq 1.
\tag{C.9}
$$

In the exact description, the density matrix of the radiation is not pure. We therefore expect a finite answer, so the relative entropy formula is not immediately applicable. It would be interesting to apply the planar resummation techniques of [20] to compute the finite answer.

C.3 Global symmetry in the ensemble

Here we would like to make some distinctions about the various kinds of symmetries in JT gravity + EoW branes.

1. The simplest symmetry is the semi-classical symmetry of the theory. This means that if we only include the trivial, disconnected topologies in the path integral, we will have a matter theory which has an exact global U(k) symmetry. In particular, the overlaps between different brane states $\langle \psi_i | \psi_j \rangle = \delta_{ij}$ are exactly invariant under $|\psi_i\rangle \to \sum_j U_{ij} |\psi_j\rangle$.

2. If we include higher topologies in the path integral, it has been shown that the path integral is dual to an ensemble of theories. The ensemble is a probability distribution over the Hamiltonian (a random matrix) and a set of pure states (k random vectors). If we consider one instance of the disorder average, the resulting theory will generically not have a U(k) global symmetry. For example, the inner products between different pure states will be slightly different.

3. Nevertheless, the ensemble has a U(k) symmetry. Note that unlike in point 1, the U(k) symmetry acts in *theory space*. It is not the symmetry group of an individual theory. Specifying each theory means specifying the values of the Hamiltonian and the pure state. The statement about the symmetry of the ensemble is that if we take such one instance of the ensemble t_1, and then act with $g \in$ U(k) on the random vectors to obtain a new theory $t_2 = g(t_1)$, the probability $P(t_2) = P(t_1)$. If we think of the ensemble as a third-quantization [61], we might say that U(k) is a symmetry in the third-quantized theory.

References

[1] S.B. Giddings and A. Strominger, *Loss of Incoherence and Determination of Coupling Constants in Quantum Gravity*, *Nucl. Phys. B* **307** (1988) 854 [INSPIRE].

[2] R. Kallosh, A.D. Linde, D.A. Linde and L. Susskind, *Gravity and global symmetries*, *Phys. Rev. D* **52** (1995) 912 [hep-th/9502069] [INSPIRE].

[3] N. Arkani-Hamed, L. Motl, A. Nicolis and C. Vafa, *The String landscape, black holes and gravity as the weakest force*, *JHEP* **06** (2007) 060 [hep-th/0601001] [INSPIRE].

[4] T. Banks and N. Seiberg, *Symmetries and Strings in Field Theory and Gravity*, *Phys. Rev. D* **83** (2011) 084019 [arXiv:1011.5120] [INSPIRE].

[5] D. Harlow and H. Ooguri, *Symmetries in quantum field theory and quantum gravity*, arXiv:1810.05338 [INSPIRE].

[6] D. Harlow and H. Ooguri, *Constraints on Symmetries from Holography*, *Phys. Rev. Lett.* **122** (2019) 191601 [arXiv:1810.05337] [INSPIRE].

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

[8] L. Susskind, *Trouble for remnants*, hep-th/9501106 [INSPIRE].

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

[10] G. Penington, *Entanglement Wedge Reconstruction and the Information Paradox*, *JHEP* **09** (2020) 002 [arXiv:1905.08255] [INSPIRE].

[11] 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].

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

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

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

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

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

[17] 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].

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

[19] A. Almheiri, T. Hartman, J. Maldacena, E. Shaghoulian and A. Tajdini, *Replica Wormholes and the Entropy of Hawking Radiation*, *JHEP* **05** (2020) 013 [arXiv:1911.12333] [INSPIRE].

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

[21] L.F. Abbott and M.B. Wise, *Wormholes and Global Symmetries*, *Nucl. Phys. B* **325** (1989) 687 [INSPIRE].

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

[23] D. Harlow and E. Shaghoulian, *Global symmetry, Euclidean gravity, and the black hole information problem*, arXiv:2010.10539 [INSPIRE].

[24] P.-S. Hsin, L.V. Iliesiu and Z. Yang, *A violation of global symmetries from replica wormholes and the fate of black hole remnants*, arXiv:2011.09444 [INSPIRE].

[25] 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].

[26] 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].

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

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

[29] R. Bousso and M. Tomašević, *Unitarity From a Smooth Horizon?*, *Phys. Rev. D* **102** (2020) 106019 [arXiv:1911.06305] [INSPIRE].

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

[31] C. Akers, N. Engelhardt and D. Harlow, *Simple holographic models of black hole evaporation*, *JHEP* **08** (2020) 032 [arXiv:1910.00972] [INSPIRE].

[32] M. Rozali, J. Sully, M. Van Raamsdonk, C. Waddell and D. Wakeham, *Information radiation in BCFT models of black holes*, *JHEP* **05** (2020) 004 [arXiv:1910.12836] [INSPIRE].

[33] F.F. Gautason, L. Schneiderbauer, W. Sybesma and L. Thorlacius, *Page Curve for an Evaporating Black Hole*, *JHEP* **05** (2020) 091 [arXiv:2004.00598] [INSPIRE].

[34] T. Anegawa and N. Iizuka, *Notes on islands in asymptotically flat 2d dilaton black holes*, *JHEP* **07** (2020) 036 [arXiv:2004.01601] [INSPIRE].

[35] T. Hartman, E. Shaghoulian and A. Strominger, *Islands in Asymptotically Flat* 2*D Gravity*, *JHEP* **07** (2020) 022 [arXiv:2004.13857] [INSPIRE].

[36] V. Balasubramanian, A. Kar, O. Parrikar, G. Sárosi and T. Ugajin, *Geometric secret sharing in a model of Hawking radiation*, *JHEP* **01** (2021) 177 [arXiv:2003.05448] [INSPIRE].

[37] K. Hashimoto, N. Iizuka and Y. Matsuo, *Islands in Schwarzschild black holes*, *JHEP* **06** (2020) 085 [arXiv:2004.05863] [INSPIRE].

[38] H. Geng and A. Karch, *Massive islands*, *JHEP* **09** (2020) 121 [arXiv:2006.02438] [INSPIRE].

[39] H.Z. Chen, R.C. Myers, D. Neuenfeld, I.A. Reyes and J. Sandor, *Quantum Extremal Islands Made Easy, Part I: Entanglement on the Brane*, *JHEP* **10** (2020) 166 [arXiv:2006.04851] [INSPIRE].

[40] Y. Chen, V. Gorbenko and J. Maldacena, *Bra-ket wormholes in gravitationally prepared states*, *JHEP* **02** (2021) 009 [arXiv:2007.16091] [INSPIRE].

[41] T. Hartman, Y. Jiang and E. Shaghoulian, *Islands in cosmology*, *JHEP* **11** (2020) 111 [arXiv:2008.01022] [INSPIRE].

[42] V. Balasubramanian, A. Kar and T. Ugajin, *Islands in de Sitter space*, *JHEP* **02** (2021) 072 [arXiv:2008.05275] [INSPIRE].

[43] W. Sybesma, *Pure de Sitter space and the island moving back in time*, arXiv:2008.07994 [INSPIRE].

[44] A. Almheiri, T. Hartman, J. Maldacena, E. Shaghoulian and A. Tajdini, *The entropy of Hawking radiation*, arXiv:2006.06872 [INSPIRE].

[45] R. Jackiw, *Lower Dimensional Gravity*, *Nucl. Phys. B* **252** (1985) 343 [INSPIRE].

[46] C. Teitelboim, *Gravitation and Hamiltonian Structure in Two Space-Time Dimensions*, *Phys. Lett. B* **126** (1983) 41 [INSPIRE].

[47] A. Almheiri and J. Polchinski, *Models of AdS$_2$ backreaction and holography*, *JHEP* **11** (2015) 014 [arXiv:1402.6334] [INSPIRE].

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

[49] H. Casini and M. Huerta, *Reduced density matrix and internal dynamics for multicomponent regions*, *Class. Quant. Grav.* **26** (2009) 185005 [arXiv:0903.5284] [INSPIRE].

[50] X. Dong and H. Wang, *Enhanced corrections near holographic entanglement transitions: a chaotic case study*, *JHEP* **11** (2020) 007 [arXiv:2006.10051] [INSPIRE].

[51] D. Marolf, S. Wang and Z. Wang, *Probing phase transitions of holographic entanglement entropy with fixed area states*, *JHEP* **12** (2020) 084 [arXiv:2006.10089] [INSPIRE].

[52] C. Akers and G. Penington, *Leading order corrections to the quantum extremal surface prescription*, arXiv:2008.03319 [INSPIRE].

[53] N. Lashkari, *Relative Entropies in Conformal Field Theory*, *Phys. Rev. Lett.* **113** (2014) 051602 [arXiv:1404.3216] [INSPIRE].

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

[55] J.M. Maldacena and L. Maoz, *Wormholes in AdS*, *JHEP* **02** (2004) 053 [hep-th/0401024] [INSPIRE].

[56] J.S. Cotler et al., *Black Holes and Random Matrices*, *JHEP* **05** (2017) 118 [*Erratum ibid.* **09** (2018) 002] [arXiv:1611.04650] [INSPIRE].

[57] P. Saad, S.H. Shenker and D. Stanford, *A semiclassical ramp in SYK and in gravity*, arXiv:1806.06840 [INSPIRE].

[58] P. Saad, S.H. Shenker and D. Stanford, *JT gravity as a matrix integral*, arXiv:1903.11115 [INSPIRE].

[59] P. Saad, *Late Time Correlation Functions, Baby Universes, and ETH in JT Gravity*, arXiv:1910.10311 [INSPIRE].

[60] J. Pollack, M. Rozali, J. Sully and D. Wakeham, *Eigenstate Thermalization and Disorder Averaging in Gravity*, *Phys. Rev. Lett.* **125** (2020) 021601 [arXiv:2002.02971] [INSPIRE].

[61] D. Marolf and H. Maxfield, *Transcending the ensemble: baby universes, spacetime wormholes, and the order and disorder of black hole information*, *JHEP* **08** (2020) 044 [arXiv:2002.08950] [INSPIRE].

[62] S.B. Giddings and G.J. Turiaci, *Wormhole calculus, replicas, and entropies*, *JHEP* **09** (2020) 194 [arXiv:2004.02900] [INSPIRE].

[63] J. McNamara and C. Vafa, *Baby Universes, Holography, and the Swampland*, arXiv:2004.06738 [INSPIRE].

[64] R. Bousso and E. Wildenhain, *Gravity/ensemble duality*, *Phys. Rev. D* **102** (2020) 066005 [arXiv:2006.16289] [INSPIRE].

[65] S.R. Coleman, *Black Holes as Red Herrings: Topological Fluctuations and the Loss of Quantum Coherence*, *Nucl. Phys. B* **307** (1988) 867 [INSPIRE].

[66] S.B. Giddings and A. Strominger, *Baby Universes, Third Quantization and the Cosmological Constant*, *Nucl. Phys. B* **321** (1989) 481 [INSPIRE].

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

[68] J. Liu, *Scrambling and decoding the charged quantum information*, *Phys. Rev. Res.* **2** (2020) 043164 [arXiv:2003.11425] [INSPIRE].

[69] J. Polchinski, *Monopoles, duality, and string theory*, *Int. J. Mod. Phys. A* **19S1** (2004) 145 [hep-th/0304042] [INSPIRE].

[70] R.E. Arias, H. Casini, M. Huerta and D. Pontello, *Entropy and modular Hamiltonian for a free chiral scalar in two intervals*, *Phys. Rev. D* **98** (2018) 125008 [arXiv:1809.00026] [INSPIRE].