Gravity/ensemble duality

Raphael Bousso[1,2,*] and Elizabeth Wildenhain[1,†]

[1]*Center for Theoretical Physics and Department of Physics, University of California, Berkeley, California 94720, USA*

[2]*Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA*

(Received 30 June 2020; accepted 1 September 2020; published 16 September 2020)

For the first time, a gravitational calculation was recently shown to yield the Page curve for the entropy of Hawking radiation, consistent with unitary evolution. However, the calculation takes as essential input Hawking's result that the radiation entropy becomes large at late times. We call this apparent contradiction the state paradox. We exhibit its manifestations in standard and doubly holographic settings, with and without an external bath. We clarify which version(s) of the Ryu-Takayanagi prescription apply in each setting. We show that the two possible homology rules in the presence of a braneworld generate a bulk dual of the state paradox. The paradox is resolved if the gravitational path integral computes averaged quantities in a suitable ensemble of unitary theories, a possibility supported independently by several recent developments.

DOI: 10.1103/PhysRevD.102.066005

I. INTRODUCTION

The black hole information paradox [1] is a conflict between quantum mechanics and general relativity. By the principle of unitarity, quantum information should be preserved in a scattering process that returns all energy to a distant observer. A pure in-state should evolve to a pure out-state. Hawking's calculation of black hole radiation [2,3], however, implies that only the energy is returned, but not the information. For decades, the only concrete evidence against information loss came from an indirect argument: assuming the AdS/CFT correspondence [4], the S-matrix can be computed in the CFT and so must be unitary.

A. Unitarity from Ryu-Takayanagi

For the first time, a purely bulk calculation was recently found to support unitarity [5,6]. This challenges Hawking's conclusion directly, rather than through an asserted duality. The new analysis does not identify an error in Refs. [2,3]; in fact, it uses Hawking's calculation. But it asks a different question, which leads to a different conclusion.

Hawking asked about the quantum state of the black hole radiation and found it to be a thermal state, $\rho_{\mathrm{Haw}}(t)$. Its von Neumann entropy,

$$S = -\mathrm{tr}\rho_{\mathrm{Haw}} \log \rho_{\mathrm{Haw}}, \tag{1.1}$$

rises monotonically as more radiation is produced. When the black hole is fully evaporated, S will be of order $\mathcal{A}_0/4G$, where $\mathcal{A}_0$ is the initial black hole area. See Fig. 1 (middle subfigure, upper graph).

By contrast, Refs. [5,6] ask only about the entropy of the radiation, not its state. The entropy S is computed not via Eq. (1.1), but as the analytic continuation of the nth Renyi entropy to $n = 1$. In the presence of gravity, this method is compactly encoded [7] in the Ryu-Takayanagi (RT) prescription [8–12]. In Sec. III, we will give a precise definition of the RT prescription in the setting of Refs. [5,6]. Schematically,

$$S(\mathrm{radiation}) = S_{\mathrm{gen}}[\mathrm{EW}(\mathrm{radiation})], \tag{1.2}$$

where EW denotes a region called the entanglement wedge, and

$$S_{\mathrm{gen}}(\mathrm{EW}) = \frac{\mathcal{A}(\partial \mathrm{EW})}{4G} + S(\mathrm{EW}) \tag{1.3}$$

is the generalized entropy. Here ∂ denotes the boundary of a region, $\mathcal{A}$ denotes the area, and G is Newton's constant.

The spacetime and its matter fields are computed using Hawking's approach, semiclassical gravity. But by using Eq. (1.2) instead of Eq. (1.1), one finds that $S(\mathrm{radiation})$ follows the "Page curve" demanded by unitary evolution.

*bousso@berkeley.edu
†elizabeth_wildenhain@berkeley.edu

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Funded by SCOAP³.

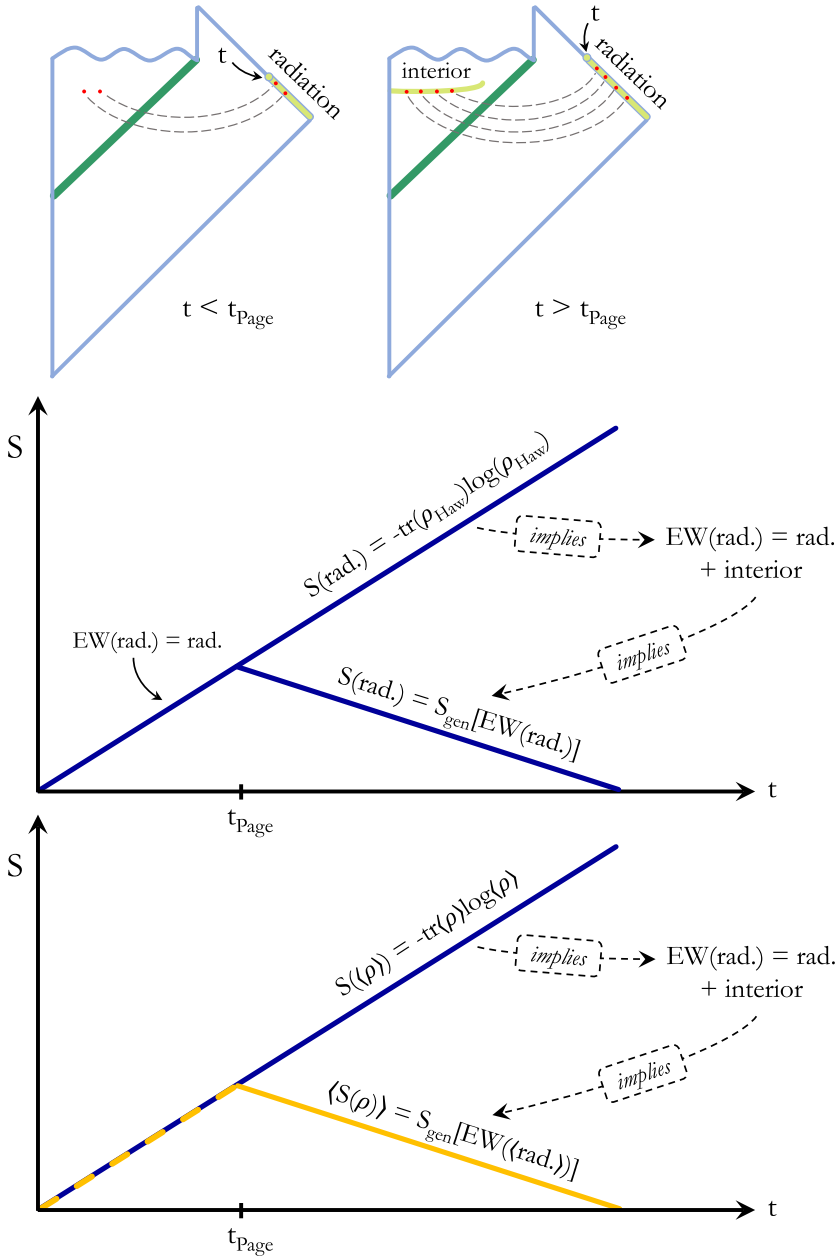

FIG. 1. Top: Penrose diagrams for an evaporating black hole. The light green region is the entanglement wedge of the radiation that has arrived at infinity before (left) and after (right) the Page time. Middle: state paradox. The RT prescription yields the Page curve for the entropy of the radiation, but only if the same entropy is assumed to follow Hawking's rising curve when determining the entanglement wedge. Bottom: resolution of the state paradox by gravity/ensemble duality. The ensemble-averaged state is mixed, and its entropy follows Hawking's curve. The ensemble-averaged entropy follows the Page curve.

It rises until the Page time, t_{Page}, when the black hole and radiation entropies are equal. Then S(radiation) falls, ultimately vanishing when the evaporation is complete. See Fig. 1 (middle subfigure, lower graph).

Before the Page time, EW(radiation) is the radiation itself (Fig. 1, left Penrose diagram). The RT prescription adds nothing new; the entropy rises because it does so in Hawking's calculation. After the Page time (right Penrose diagram), a minimality condition in the definition of the entanglement wedge implies[1] that EW(radiation) contains both the radiation and a disconnected "island," the black hole interior:

[1]When the minimality condition results in an island, it has been called the "island rule" [13]. However, this is not a new rule nor a modification of RT. The existence of islands after the Page time already follows from the RT prescription in the final form given to it by Engelhardt and Wall [12].

$$\mathrm{EW}(\mathrm{radiation}) = \mathrm{radiation} + \mathrm{black\ hole\ interior} \quad (t > t_{\mathrm{Page}}). \tag{1.4}$$

In Hawking's analysis, the interior together with the Hawking radiation are in a pure state. Hence the von Neumann entropy $S(\mathrm{EW})$ vanishes, and only the area of the boundary of the island contributes. This boundary is approximately the black hole horizon, so

$$S_{\mathrm{gen}}[\mathrm{EW}(\mathrm{radiation})] = \frac{\mathcal{A}(\mathrm{horizon})}{4G} \quad (t > t_{\mathrm{Page}}). \tag{1.5}$$

The horizon area decreases as the black hole evaporates, yielding the falling part of the Page curve.

B. State paradox

The breakthrough of Refs. [5,6] involves an apparent paradox: it makes use of Hawking's result that $S(\mathrm{radiation})$ increases monotonically for all times, in order to reach the final conclusion that it does not. Through Eq. (1.5), the radiation appears on both sides of Eq. (1.2). Hawking's Eq. (1.1) is invoked in evaluating its entropy on the right-hand side (RHS) of Eq. (1.2). Thus on the RHS, $S(\mathrm{radiation})$ (without the island) follows Hawking's monotonically increasing curve. This is a crucial ingredient, because it triggers the inclusion of the black hole interior in EW(radiation) after the Page time. On the LHS, $S(\mathrm{radiation})$ then follows the Page curve.

This is a contradiction. The S-matrix is an observable, so the state of the Hawking radiation cannot be ambiguous. Therefore, its von Neumann entropy cannot have two different values.[2] We will call this contradiction the *state paradox*.

One possible resolution of the paradox is that the RT prescription is an uncontrolled approximation. Our confidence in the RT prescription derives from its success in the context of AdS/CFT, where the CFT entropy can often be independently computed and shown to agree. However, under certain assumptions, the RT prescription follows directly from a bulk path integral computation [7], evaluated in the saddle point approximation. It is obtained as the analytic continuation to $n = 1$ of the n-th Renyi entropies of the radiation, which can be computed from a path integral using the replica trick. After the Page time, one finds that the dominant saddle point has wormholes connecting the replicas [15,16]; see Ref. [17] for a pedagogical review.

Thus the RT prescription has nothing to do with AdS/CFT; the nonperturbative completeness of the CFT is not used. RT can be applied even in asymptotically flat space, for example to compute the entropy of radiation that has arrived at the conformal boundary [18,19]. RT is an advanced analogue of the Euclidean computation of the thermodynamic entropy of a black hole by Gibbons and Hawking [20]. It cleverly extracts information about the full quantum gravity theory from a path integral approximation.

This is not a controlled approximation. It need not agree with the full quantum gravity theory, and when it does, it need not be self-consistent. This could explain the state paradox: perhaps Eq. (1.2) just happens to compute the correct statistical entropy from the incorrect state. (See Ref. [21] for a discussion of related ideas.) And one day, perhaps, an even more sophisticated application of the Euclidean gravity path integral will be shown to yield the correct state of the Hawking radiation.

C. Gravity/ensemble duality

A different, intriguing possibility is that there exists a novel kind of duality: between an appropriately defined version of the gravitational path integral, and an ensemble of quantum mechanical theories without gravity. This can resolve the state paradox [22]. According to this proposal, $S(\mathrm{radiation})$ takes two different values on the two sides of Eq. (1.2) because it is not the same quantity on the two sides.

On the left side, it is the ensemble average of the entropy, so we should replace $S(\mathrm{radiation}) \to \langle S(\rho)\rangle$. See Fig. 1 (bottom subfigure, lower graph). On the right side, the entanglement wedge is determined from the entropy of the ensemble-averaged state of the radiation, $S(\langle\rho\rangle)$. See Fig. 1 (bottom subfigure, upper graph). Because the von Neumann entropy is not a linear function of the state, generically $\langle S(\rho)\rangle \neq S(\langle\rho\rangle)$.

We now describe this proposal in more detail. Let ν label unitary theories, each capable of computing a pure Hawking radiation out-state from any pure in-state. Let

$$\langle x\rangle \equiv \int d\nu c(\nu) x^{(\nu)} \tag{1.6}$$

denote an appropriately weighted average of the quantity x computed in the different theories. Let ρ_{in} be the initial state before the black hole forms, and $\rho_{\mathrm{out}}^{(\nu)}$ be the final state of the radiation when the black hole has fully evaporated. Since each theory is unitary, we have

[2] One might be tempted to declare that S computed from Eq. (1.1) is only a coarse-grained entropy (even though no coarse-graining is manifest in Hawkings calculation). But the second term on the right side of Eq. (1.3) is a fine-grained von Neumann entropy, and it is this fine-grained entropy that determines EW (radiation) in Eq. (1.2). The island that leads to the Page curve can only be included if the *fine-grained* entropy of the radiation continues to grow after the Page time. This is achieved by taking Hawking's calculation seriously at this step in the calculation, as a fine-grained entropy. Moreover, if the Page curve was assumed from the beginning, then the smooth horizon shown in the top diagrams in Fig. 1 would be inconsistent [14] and so cannot enter the analysis at all. Finally, rejecting Eq. (1.1) as a fine-grained entropy would amount to putting in the Page curve by hand. With the Page curve for the radiation as input, Eq. (1.2) would reproduce the Page curve trivially as an identity, not by inclusion of an island.

$$S(\rho_{\text{out}}^{(\nu)}) = 0 \quad \text{for all } \nu, \tag{1.7}$$

and hence

$$\langle S(\rho_{\text{out}})\rangle = 0. \tag{1.8}$$

But in general, the final states $\rho_{\text{out}}^{(\nu)}$ will be different in different theories. We now assume that their ensemble average is the thermal state predicted by Hawking:

$$\langle \rho_{\text{out}}\rangle = \rho_{\text{Haw}}. \tag{1.9}$$

With these assumptions, Hawking's calculation computes the averaged out-state $\langle\rho_{\text{out}}\rangle$; and in the same spacetime, the RT prescription correctly computes the averaged entropy:

$$\langle S[\rho_{\text{out}}]\rangle = S_{\text{gen}}[\text{EW}(\langle\rho_{\text{out}}\rangle)]. \tag{1.10}$$

Moreover, this holds at all times. Let $\rho(t) = \text{tr}_{>t}\,\rho_{\text{out}}$ be the state of the radiation subsystem that has escaped to a distant region by the time t. The ensemble version of the RT prescription, Eq. (1.2) states that

$$\langle S[\rho(t)]\rangle = S_{\text{gen}}[\text{EW}(\langle\rho(t)\rangle)]. \tag{1.11}$$

No contradiction arises. The ensemble average of the entropy will follow the Page curve, while the entropy of the ensemble average follows Hawking's curve.

The state paradox and its resolution by gravity/ensemble duality was first described in a slightly different setting [22], which we will review in Sec. II. Another compelling argument for gravity/ensemble duality comes from the fact that the partition function on multiple copies of a boundary need not factorize when it is computed from a bulk gravity dual, because connected geometries can contribute [15,16]. It would be interesting to understand the detailed relation between these arguments.

The duality between JT gravity [23,24] and a random matrix ensemble furnishes an important concrete example of gravity/ensemble duality [25–29]. Recently, an average over certain two-dimensional CFTs was shown to exhibit properties of an exotic three-dimensional gravity theory [30,31]. Conversely, starting with three-dimensional Einstein gravity, properties of a putative ensemble dual have been explored [32]; see also [33,34].

An ensemble of theories satisfying Eqs. (1.7) and (1.9) may not exist in all cases where the RT prescription can be applied. If it does not, then the state paradox remains unresolved. For example, type IIB supergravity on $\text{AdS}_5 \times \mathbf{S}^5$ is dual to a specific CFT [4], and no other boundary theories are presently known that have the same bulk as a coarse-grained description. If none exist, the gravitational path integral may still be expected to compute quantities that *would* be self-averaging if an ensemble *did* exist [26]. It would determine the entropy $S(\rho_{\text{out}})$ but not the state ρ_{out}.

D. Outline

In this paper, we consider several distinct settings in which the state paradox appears, and we discuss its possible resolution by gauge/ensemble duality in each case. Multiple versions of the RT prescription will apply, and we will clarify their relation.

In Sec. II, we use the RT prescription in an AdS_d bulk spacetime to derive the Page curve in the dual CFT_{d-1} [22]. The setting is distinct from that of Refs. [5,6] in that there is no external bath or auxiliary system. The radiation remains in an AdS bulk and appears only on the right side of Eq. (1.2). The left-hand side corresponds to the entropy of the CFT dual, for which a Page curve is obtained. The state paradox then arises in the CFT. The CFT entropy can also be computed as the von Neumann entropy of a CFT state constructed by applying the standard AdS/CFT extrapolate dictionary to the bulk. With this method, one finds that the CFT entropy should grow monotonically. These results are consistent only if the CFT is actually an ensemble of CFTs.

In Sec. III, we turn to the setting of Refs. [5,6]. The gravitating spacetime is coupled to an auxiliary system without gravity, into which the Hawking radiation escapes. The entropy of the auxiliary system is computed using RT. The RT prescription must first be extended so that it applies to auxiliary systems; this was initially viewed as a weak link in the analysis [6]. We argue that the correct prescription is fully determined by consistency with the setting of Sec. II. The radiation appears on both sides of Eq. (1.2), leading to the state paradox unless gravity/ensemble duality is invoked.

Several works [13,35–39] have computed the Page curve using the entanglement wedge in a "doubly holographic" dual. The fundamental object in this case is the auxiliary system containing the radiation: a "Boundary" conformal field theory or BCFT (in the sense of Refs. [40,41]), with an apparently different RT prescription [42,43]. The state paradox is somewhat obscured in this approach. To exhibit it, we deconstruct the RT prescription for BCFTs as a repeated application of the original RT prescription.

For the sake of clarity, we first develop an RT prescription for a doubly holographic setting *without* auxiliary system, in Sec. IV. In this case, the fundamental object is a regular CFT_{d-1} dual to an AdS_d bulk. The bulk matter sector is assumed to consist of a holographic CFT_d coupled to gravity. Then there exists a second holographic dual with $d+1$ dimensions. The original RT prescription computes the von Neumann entropy of a CFT_{d-1} region as the generalized entropy of its entanglement wedge in the AdS_d bulk. An adaptation of the RT prescription to braneworlds [44,45] can be used to compute generalized entropy in the AdS_d bulk using the $d+1$ dimensional bulk. We show that these steps can be combined into a one-step "squared RT" prescription for computing CFT_{d-1} entropy from a "squared entanglement wedge," EW^2, in the $d+1$ dimensional bulk.

In Sec. V, we combine the settings of the previous two sections. We consider a doubly holographic CFT_{d-1}, coupled to a (singly) holographic CFT_d that plays the role of the auxiliary system of Sec. III. In the second holographic dual, the CFT_d is part of the conformal boundary of the $d+1$ dimensional bulk. Like in Sec. IV, we show that the RT prescriptions for each holographic layer can be combined into a (one-step) squared RT prescription that uses the $d+1$ bulk to compute the von Neumann entropy of any union of subregions of the above top-level CFT_{d-1} and CFT_d. Our squared RT prescription agrees with the known RT prescription for BCFTs [42,43].

It follows that [42,43] can be deconstructed as two applications of the RT prescription. This allows us to shed light on a number of puzzling features in Refs. [13,35–39]. We find that the state paradox arises at the first step, for the Hawking radiation that has escaped to the "auxiliary" CFT_d. At this level the paradox can be resolved by replacing (at least) the CFT_{d-1} with an ensemble of such theories.

The second level of holography furnishes a bulk dual of the original state paradox. The RT prescription for braneworlds computes the entropy of subregions of the first holographic dual, in terms of bulk quantities in the second dual. Choosing the subregion to be just the radiation region, this reproduces Hawking's rising curve; choosing it to include the island as well, one again obtains the Page curve.

Finally, we observe that when the entropy of a top level CFT_d region is computed directly using the squared RT prescription [42,43], no paradox is manifest, because the $d+1$ bulk dual does not contain the radiation.

E. Discussion

The discovery of entanglement islands [5,6] provides evidence for unitarity, independently of AdS/CFT. It marks a new era in which the Page curve can be derived from gravitational physics directly. It provides independent evidence for unitary evolution. However, it does not resolve the critical question of how the information gets out.

If we insist that information is preserved when a black hole evaporates, then effective field theory or general relativity must break down substantially, at or outside of the horizon [14], at late times but while the horizon is still weakly curved. This formulation of the information paradox is called the firewall paradox.

The firewall argument suggests that Hawking's "mistake" was the perfectly reasonable assumption that the horizon of a large old black hole is smooth. The AdS/CFT correspondence can be used to strengthen this argument [46], but it has shed no light on the bulk dynamics that would produce a firewall. A number of interesting proposals attempt to reconcile unitarity with a smooth horizon; see Refs. [47,48] for a critical review. These proposals remain incomplete, and they appear to necessitate an element of nonlinearity that conflicts with the principles of quantum mechanics no less than information loss would [49–51].

The bulk path integral derivation of the Page curve has been interpreted as a resolution of the firewall paradox [5,13,16,17]. This seems plausible, since the bulk geometry involved in the calculation of the Page curve (top of Fig. 1) has a manifestly smooth horizon. However, this picture just trades the firewall paradox for the state paradox [22]. Then the question becomes how the state paradox is resolved. We see two possibilities.

Suppose that the state paradox is resolved by gravity/ensemble duality. The firewall argument [14] does not apply to the ensemble averaged state, since its evolution is not unitary. Therefore it is consistent for the horizon to be smooth. However, fundamentally it makes no sense for Nature to be described by an ensemble of unitary theories; we can just measure the couplings and then work with the one correct theory. Moreover, we do expect the unique theory describing black hole formation and evaporation—the one that applies to an experiment conducted in a lab—to preserve information. The ensemble will be useful only for computing self-averaging quantities of the correct unitary theory, since these are the same in each theory; these evidently include the entropy, but not the final state. Hence the true S-matrix must be computed from a single unitary theory, not from an ensemble. In this theory, the firewall argument still applies.

If instead there is no gravity/ensemble duality (for example, in settings where no suitable ensemble exists, or where the gravity path integral cannot be rigorously defined), then the bulk path integral (or the saddlepoint approximation to it) would have to be viewed as an uncontrolled approximation. The path integral succeeds at computing certain quantities of a single unitary boundary theory (like the entropy) but not others (like the details of the late time state). Then there is no reason to trust the smooth geometry that appears in the input to the RT calculation, any more than we should trust the large entropy of the Hawking radiation that is manifest at this step. If we believe the output of the RT calculation—the Page curve—then the firewall paradox precludes a smooth horizon.

F. Notation and conventions

A subscript on a geometric object generally indicates not *its* dimension, but the dimension of the (physical) spacetime in which the object is naturally defined. For example, M_d will denote a d-dimensional spacetime, R_d a $d-1$ dimensional spatial region in M_d, and γ_d a $d-2$ dimensional extremal surface in M_d. It is often useful to conformally rescale a manifold M_d so that a boundary can be added to it; the result is called an unphysical manifold or Penrose diagram, $\tilde{M}_d \supset \partial\tilde{M}_d$. Note that the boundary of the physical manifold, $\partial M_d \subset \partial\tilde{M}_d$, need not be empty; it consists of braneworlds or end-of-the-world (EOW) branes.

In this paper, $\langle x \rangle$ always denotes the ensemble average of x in the sense of Eq. (1.6). Angular brackets never denote a quantum expectation value.

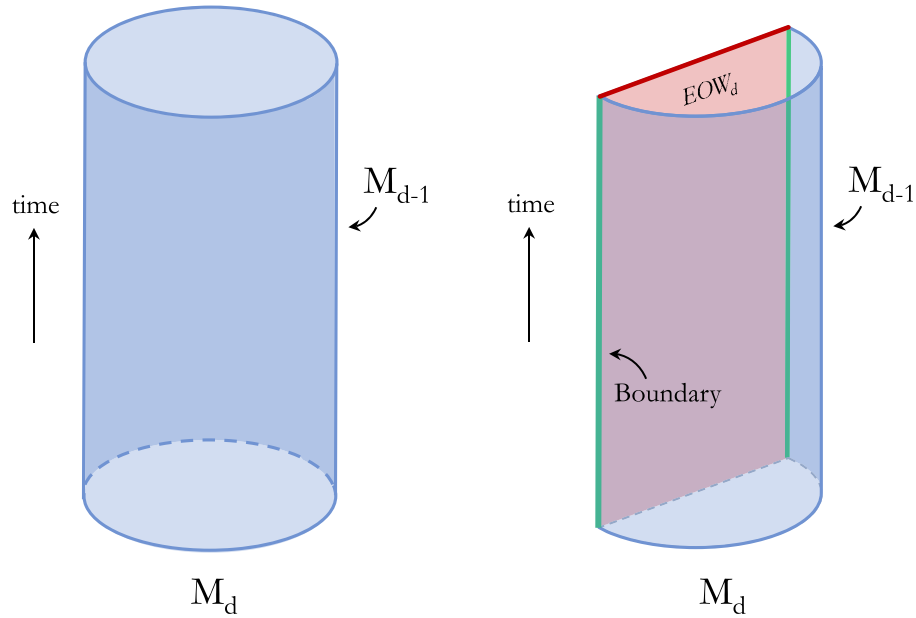

FIG. 2. Examples of holographic duality. Left: the solid bulk M_d is dual to a holographic CFT_{d-1} on M_{d-1} (blue boundary). Right: in this example, M_{d-1} is a manifold with boundary, so the boundary theory is a BCFT_{d-1} and M_d contains an end of the world brane EOW_d. (Despite the appearance of a BCFT this is a "singly holographic" example. In Secs. IV and V we will consider a doubly holographic setting where the EOW_d is a braneworld that localizes gravity and contains a holographic CFT_d.)

The term *Boundary conformal field theory* (BCFT) refers to the fact that such a theory lives on a manifold *with* Boundary, not to the fact that it lives on the conformal boundary *of* some AdS spacetime. We will capitalize "Boundary" whenever it is used in the sense of a BCFT. For example, "boundary entropy" might refer to the von Neumann entropy of a CFT region on the conformal boundary $M_{d-1} = \partial\tilde{M}_d$, whereas "Boundary entropy" is a specific BCFT parameter defined by Cardy [40].

Throughout this paper we assume $d > 2$ for convenience. The case $d = 2$ would frequently require a special treatment; see for example Eq. (4.3). This would clutter the presentation. However, the qualitative aspects of our analysis apply in $d = 2$, and hence to the many recent works that studied entanglement islands in JT gravity and other two-dimensional models, such as Refs. [6,13]. Related to this choice, in examples involving braneworlds we only consider induced gravity on the brane (i.e., the localized graviton due to embedding of the brane in AdS [52]). We never add an additional gravitational action on the brane, because in $d > 2$ this is not necessary.

II. GRAVITY/ENSEMBLE DUALITY WITHOUT A BATH

In this section, we exhibit a version of the state paradox in which only the standard RT prescription is needed [22]. There is no auxiliary system or bath, and there is only one layer of holography.

A. General setup

Consider a $d-1$ dimensional[3] holographic conformal field theory CFT_{d-1} with central charge c_{d-1}, living on a manifold M_{d-1}; see Fig. 2. Its bulk dual will be an asymptotically AdS_d spacetime M_d,[4] such that the unphysical spacetime (or Penrose diagram) conformally related to M_d [53] is

$$\tilde{M}_d = M_d \cup M_{d-1}; \tag{2.1}$$

thus M_{d-1} is the conformal boundary of M_d. The AdS_d curvature length L_d is related to the central charge by

$$\frac{L_d^{d-2}}{G_d} \sim c_{d-1}, \tag{2.2}$$

where $G_d = \ell_d^{d-2}$ is Newton's constant in the d-dimensional bulk.

We shall denote a standard holographic duality of this type as follows:

[3]For consistency with the later sections on double holography, we deviate here from the usual convention of using d for the boundary spacetime dimension.

[4]In general this spacetime can contain additional factors, e.g., $\mathrm{AdS}_d \times \mathbf{S}^{d'}$, so it need not actually be d-dimensional. In order to keep the discussion simple, we will assume that it is; generalizations are straightforward.

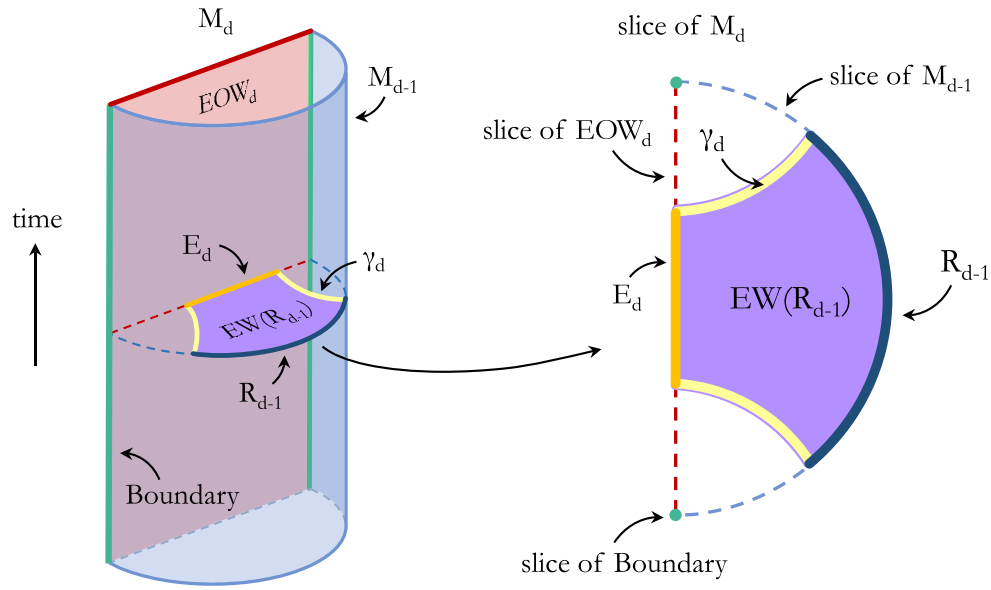

FIG. 3. RT prescription, applied in the setting shown on the right of Fig. 2. The entropy of the boundary region R_{d-1} is given by the generalized entropy of its entanglement wedge $\mathrm{EW}(R_{d-1})$. γ_d is the quantum extremal surface.

$$M_{d-1} \to M_d, \tag{2.3}$$

where the arrow reminds us that in general, this duality is not truly an equivalence. Rather, the lower dimensional field theory without gravity can be viewed as the nonperturbative completion of the bulk theory.

Note that M_{d-1} may itself have a boundary, as in Fig. 2. The spacetime M_d may also be a manifold with boundary [53], commonly referred to as an "end of the world brane" or EOW:

$$\mathrm{EOW}_d \equiv \partial M_d. \tag{2.4}$$

In particular, if M_{d-1} is a manifold with boundary, then the CFT_{d-1} is a "Boundary conformal field theory" (BCFT),[5] and the bulk M_d will contain an $\mathrm{EOW}_d \neq \varnothing$ anchored on the Boundary ∂M_{d-1} [42,43]. An EOW can also exist in settings where the Lorentzian CFT_{d-1} has no Boundary [54,55]. They must be included in the gravitational path integral.

B. Ryu-Takayanagi prescription

We now formulate the holographic prescription for computing the von Neumann entropy of a boundary region from bulk quantities. This was first proposed by Ryu and Takayanagi [8,9] for stationary states. It was generalized to the time-dependent case by Hubeny, Rangamani, and Takayanagi [10], and to the BCFT case by Takayanagi and collaborators [42,43]. A quantum-corrected prescription was first proposed by Faulkner, Lewkowycz and Maldacena [11]. It was extended to all orders by Engelhardt and Wall [56], whose elegant formulation highlights the central role of generalized entropy. This final formulation is essential for the existence of islands,[6] and it is the only one we will review here. We will refer to it as the RT prescription for short, with apologies to all others involved in its development. We aim to make it clear throughout this paper that islands are part and parcel of this prescription. They do not constitute a new ingredient, but a long-overlooked consequence. The recent recognition of their existence [5,6] has been profoundly impactful.

Let $R_{d-1} \subset M_{d-1}$ be an achronal region (see Fig. 3).[7] We can think of R_{d-1} as a subregion at some instant of time, to which the CFT_{d-1} state may be restricted. The von Neumann entropy $S(R_{d-1})$ of the restricted CFT_{d-1} state is given by the generalized entropy of its entanglement wedge,

$$S(R_{d-1}) = S_{\mathrm{gen}}[\mathrm{EW}(R_{d-1})]. \tag{2.5}$$

The generalized entropy $S_{\mathrm{gen}}(X_d)$ [57] of an arbitrary achronal region $X_d \subset M_d$ is the sum of its gravitational entropy and the von Neumann entropy S of the quantum fields in the region X_d:

$$S_{\mathrm{gen}}(X_d) = \frac{\mathcal{A}(\partial X_d)}{4G_d} + S(X_d). \tag{2.6}$$

[5]See the notation section at the end of the Introduction.

[6]Because the empty surface always has less classical area than the boundary of an island, area minimization cannot lead to an island. It is vital that the generalized entropy is minimized.

[7]An achronal region is a submanifold of codimension 1 (in the spacetime) which contains no two points connected by a timelike curve.

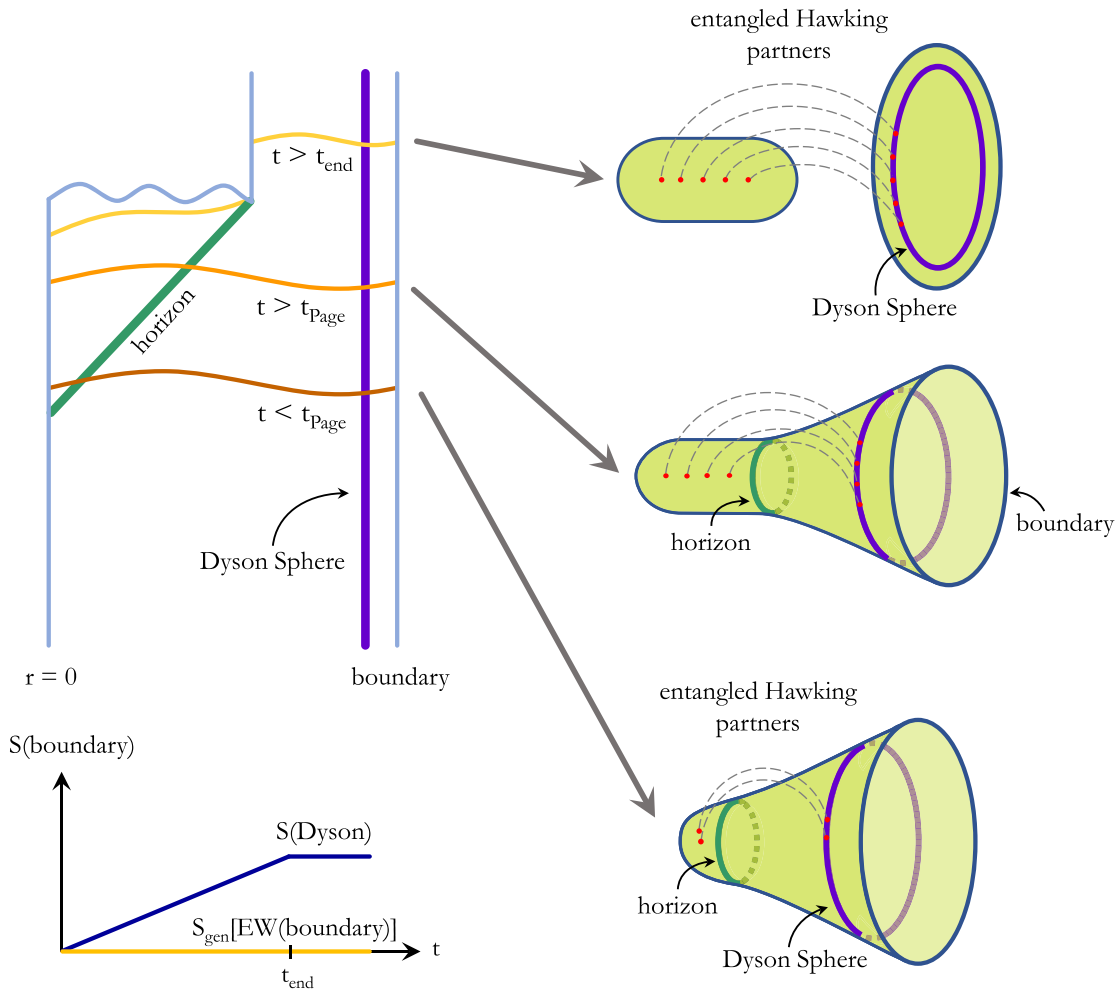

FIG. 4. Hawking radiation is absorbed by a distant Dyson sphere near the boundary. In Hawking's semiclassical analysis, the Dyson sphere entropy will grow monotonically. The quantum state on the global bulk slices shown is pure. Each global slice is the entanglement wedge of its respective boundary slices. Thus the RT prescription implies that the entropy of the global boundary vanishes, as required by CFT unitarity. However, at late times the extrapolate dictionary demands that $S(\text{boundary}) = S(\text{Dyson})$. This contradiction is the state paradox.

Here $\mathcal{A}(\partial X_d)$ is the area of the boundary of X_d in M_d, and G_d is Newton's constant in M_d.

The *entanglement wedge* $\mathrm{EW}(R_{d-1})$ is an achronal region X_d[8] in M_d, that satisfies the following conditions:

(1) *Homology*: $\partial X_d = \gamma_d \cup R_{d-1} \cup E_d$, where $\gamma_d \subset M_d - \mathrm{EOW}_d$, and $E_d \subset \mathrm{EOW}_d$.[9] See Fig. 3.

(2) *Stationarity*: $S_{\mathrm{gen}}(X_d)$ is stationary under variations of γ_d.

(3) *Minimality*: X_d is has the smallest S_{gen} among all regions with the above properties.

A surface γ_d satisfying the homology constraint (1) and the stationarity condition (2) is called *quantum extremal*[10] with respect to R_{d-1}. If the minimality condition (3) is also satisfied, then γ_d is called the *RT surface* of R_{d-1}. Note that γ_d may be the empty set; for example, see Fig. 4 below. Also, γ_d may contain disconnected components that end neither on R_{d-1} nor on E; for example, see Fig. 5 below.

C. Simple boundary unitarity from a semiclassical bulk

It was recently shown that the RT prescription applied to semiclassical bulk evolution yields an entropy consistent with boundary unitarity [5,6], for Hawking radiation extracted into an auxiliary system. This argument requires

[8]One can also define $\mathrm{EW}(R)$ to be the (d-dimensional) domain of dependence of this region. Since all Cauchy slices of the domain of dependence have the same generalized entropy, we will use these definitions interchangeably.

[9]Strictly, this is a statement about the image of X_d in the unphysical spacetime $\tilde{M}_d$.

[10]This is conventional. "Quantum stationary" would be more appropriate terminology, as the generalized entropy can be both increased and decreased at second order by suitable deformations.

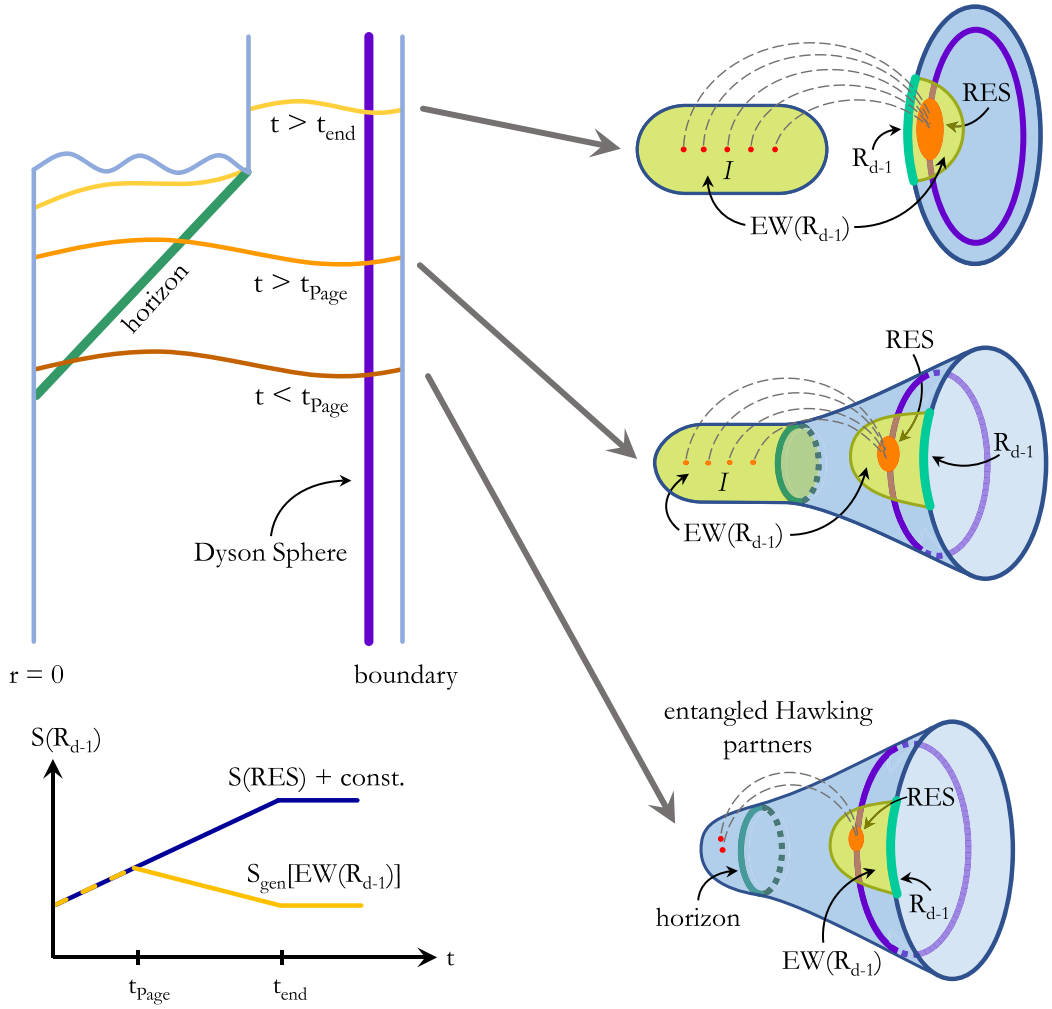

FIG. 5. Compared to Fig. 4, the Hawking radiation is collected in a localized reservoir on the Dyson sphere. The RT prescription is applied to a nearby boundary region R_{d-1}. The entanglement wedge $\mathrm{EW}(R_{d-1})$ is shown in light green. After the Page time, it contains a disconnected island I, the black hole interior, because this choice minimizes the generalized entropy. This yields the Page curve for $S(R_{d-1})$. However, the extrapolate dictionary would yield Hawking's curve; this is the state paradox.

an extension of the RT prescription that includes auxiliary systems. We will show in Sec. III A that this extension is uniquely determined by physical considerations.

However, the main result of Refs. [5,6] can be obtained without involving an auxiliary system, using only the standard RT prescription, Eq. (2.5). Here we summarize this argument; further details are discussed in Ref. [22].

Consider a CFT_{d-1} on $M_{d-1} = \mathbf{S}^{d-2} \times \mathbf{R}$. In the vacuum, the gravity dual M_d would be global AdS_d. However, we shall take M_d to be a black hole formed from collapse of matter in a pure quantum state. The black hole is surrounded by a distant detector sphere ("Dyson sphere"), initially in some pure reference state. By the extrapolate dictionary, the initial boundary state must be pure. As the black hole evaporates, the Dyson sphere absorbs all of the Hawking radiation (see Fig. 4).

Let $\Sigma_{d-1}(t)$ be a family of Cauchy surfaces (time slices) of the boundary M_{d-1}. Each such slice will be a sphere $\mathbf{S}^{d-2}$. Three slices are shown in Fig. 4. We will apply the RT prescription to every slice, but first it will be useful to make some further definitions. Let $\Sigma_d(t)$ be a Cauchy surface of $M_d(t)$ bounded by $\Sigma_{d-1}(t)$. (In $\tilde{M}_d$, $\Sigma_{d-1} = \partial\Sigma_d$.) For boundary slices that lie in the future of the endpoint of the evaporation process, we define $\Sigma_d(t)$ to include a disconnected component, a Cauchy slice of the black hole interior (see Fig. 4, yellow slice at top). This can be chosen far enough from the singularity so that semiclassical gravity is applicable everywhere but in the neighborhood of the endpoint [58].

The key observation is that the entanglement wedge is the entire bulk:

$$\Sigma_d(t) = \mathrm{EW}[\Sigma_{d-1}(t)], \tag{2.7}$$

for all t. To see this, note that the homology condition is satisfied, with $\gamma_d = \varnothing$. The stationarity condition is

satisfied because no variations of γ_d exist. The minimality condition is satisfied because

$$S_{\rm gen}[\Sigma_d(t)] = 0 \tag{2.8}$$

for all t, and the generalized entropy cannot be negative.

(Strictly, one could question all three of these statements due to the breakdown of the semiclassical description at the evaporation endpoint. We assume that this small region does not contribute significant effects that invalidate our treatment of the post-evaporation entanglement wedge. In any case, the essence of our discussion requires us only to go past the Page time, but not close to or beyond the endpoint of evaporation.)

It is important to understand why Eq. (2.8) holds. The area term in Eq. (2.6) vanishes since $\gamma_d = \varnothing$. The von Neumann entropy of the matter fields vanishes because the initial bulk state is pure, and the semiclassical bulk evolution of the *global* bulk state is unitary. (Information is lost to an observer outside the black hole in this description [2], but globally the state remains pure. The interior Hawking partners and the exterior Hawking radiation together form a pure state, the vacuum at the horizon.)

By Eq. (2.5), it follows that

$$S[\Sigma_{d-1}(t)] = 0 \tag{2.9}$$

for all t. The RT prescription "predicts" that the entropy of the boundary theory vanishes at all times. Of course, this is exactly what is expected from the unitarity of the boundary CFT_{d-1}. But it is remarkable that this result is reproduced by performing a semiclassical analysis in the bulk—the same calculation that led Hawking to conclude that information is lost to bulk observers outside the black hole. This fact was perhaps not widely appreciated prior to the recent work [5,6] that derives the entire Page curve, even though it has the same import and is simpler to obtain.

D. Island and Page curve

The previous subsection explained how the RT prescription yields the vanishing global boundary entropy consistent with unitarity, despite using Hawking's semiclassical evolution in the bulk. In this subsection, we introduce a refined scenario, such that the RT prescription yields the Page curve for two complementary subsystems, the Hawking radiation and the remaining black hole. In order to implement this without introducing an external bath or auxiliary system, any absorbed Hawking radiation is immediately transferred to a localized reservoir RES taking up a small solid angle on the Dyson sphere, without loss of quantum coherence (see Fig. 5) [22].

Gravitational backreaction in the asymptotic region can be kept arbitrarily small, so the shape of any stationary surface anchored to a small boundary region R_{d-1} will be similar to that in the vacuum. We take $R_{d-1}(t) \subset \Sigma_{d-1}(t)$ to be at the same angular position as the reservoir, and just large enough so that $\mathrm{EW}[R_{d-1}(t)]$ will barely contain the reservoir (see Fig. 5). Before the Page time, the entanglement wedge has only one connected component, and we find

$$S[R_{d-1}(t)] = S_{\rm gen}[\mathrm{EW}(R_{d-1}(t))] = \frac{\mathcal{A}[\gamma_d^{\rm conn}]}{4G} + S_{\rm RES}(t). \tag{2.10}$$

The superscript refers to the fact that $\gamma_d = \gamma_d^{\rm conn}$ is connected to R_{d-1} before the Page time. By moving around ballast on the Dyson sphere, one can arrange for the asymptotic geometry in an open neighborhood of $\gamma_d^{\rm conn}$, and hence for $\mathcal{A}[\gamma_d^{\rm conn}]$ to remain fixed [22]. The entropy of the reservoir $S_{\rm RES}(t)$, however, increases as more radiation arrives. This yields the rising part of the Page curve shown in Fig. 5.

The entropy of the Dyson sphere, and of $S_{\rm RES}$ in particular, increases monotonically even after the Page time. Its state is always purified by the "Hawking partners" in the black hole interior. Inclusion of the black hole interior in the entanglement wedge will entirely wipe out the contribution $S_{\rm RES}$ to $S_{\rm gen}(R_{d-1})$ at a cost of increasing the area term by the area of the black hole. This preserves the homology condition, since it merely adds an extra component to γ_d. By its very definition, this choice becomes favorable at the Page time, when the black hole and radiation entropy are equal.

After the Page time, the minimality condition thus requires that $\mathrm{EW}(R_{d-1})$ contains a second, disconnected component I (see Fig. 5). This is called an island, in the terminology of Ref. [13]. The island is the black hole interior, bounded by a disconnected component $\gamma_d^{\rm island}(t)$ that nearly coincides with the horizon.[11] The interior of $\gamma_d^{\rm island}$ purifies the Hawking radiation, so the entropy of the reservoir no longer contributes, and $S_{\rm gen}[\mathrm{EW}(R_{d-1})]$ is given just by the area of the RT surface $\gamma_d = \gamma_d^{\rm conn} \cup \gamma_d^{\rm island}$:

$$S_{\rm gen}[\mathrm{EW}(R_{d-1})] = \frac{\mathcal{A}[\gamma_d^{\rm conn}]}{4G} + \frac{\mathcal{A}[\gamma_d^{\rm island}(t)]}{4G}. \tag{2.11}$$

The first term remains constant. But $\gamma_d^{\rm island}(t)$ shrinks with the black hole horizon as the black hole evaporates, yielding the decreasing part of the Page curve.

Thus, in the refined scenario, the RT prescription (i.e., a bulk path integral that computes the entropy) yields the Page curve for the boundary region R_{d-1}. It rises during the first half of the evaporation process, then decreases. Again,

[11]The precise location of $\gamma_d^{\rm island}$ is determined by the stationarity condition. It sits about a Planck length inside the horizon. Temporally, $\gamma_d^{\rm island}(t)$ is located at $t - t_{\rm scr}$, where $t_{\rm scr} \sim \beta \ln(\mathcal{A}[\gamma_d^{\rm island}]/4G)$ and β is the inverse temperature of the black hole [5,6].

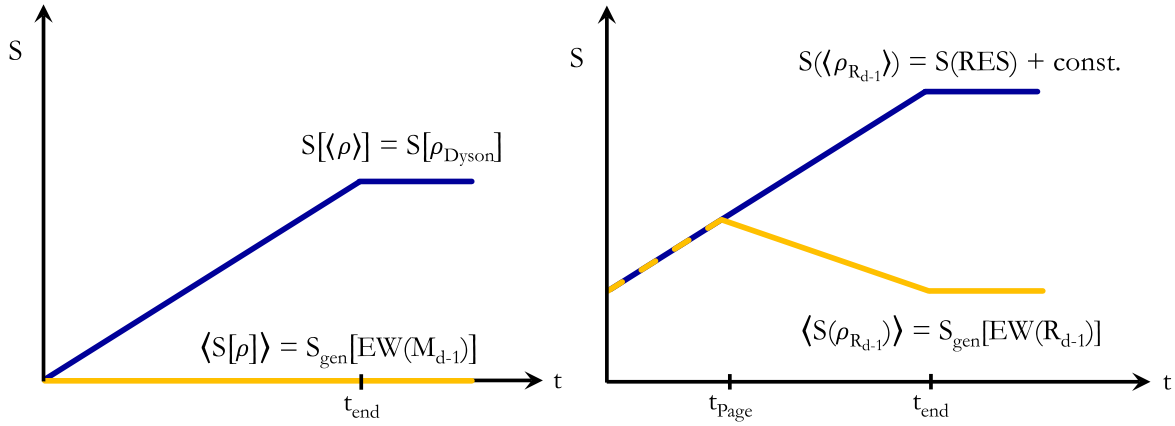

FIG. 6. Here we assume the gravity/ensemble interpretation, in the examples studied in Sec. II C (left) and Sec. II D (right). This resolves the state paradox. The RT prescription (yellow) computes the ensemble averaged entropy $\langle S\rangle$ of the full boundary (left) or of R_{d-1} (right). The extrapolate dictionary (blue) yields the average state of the ensemble, $\langle\rho\rangle$, in these regions.

this is consistent with our expectations from boundary unitarity. Entanglement wedge complementarity is manifest in the present setting, so a Page curve is also obtained for the complementary boundary region $\bar{R}_{d-1}$.

E. State paradox and ensemble interpretation

The large entropy of the Dyson sphere at late times leads to the state paradox. After the evaporation is complete, all of the (conserved) mass is in the Dyson sphere. The standard AdS/CFT dictionary can be used to construct the boundary state from the mixed state of the Dyson sphere [59]. It dictates that the boundary state must have the same entropy as the Dyson sphere. Energetic arguments preclude purification of this state by some nonlocal CFT excitations [22]. The entropy of the CFT_{d-1} should therefore grow monotonically throughout the evaporation process. But this contradicts both the RT result and the expected unitarity of the boundary theory.

We stress again that one cannot dismiss the large Dyson sphere entropy as an artifact of the semiclassical approximation, without discarding the entire RT calculation. If the reservoir RES did not have large entropy after the Page time, the black hole interior could not purify it. Then there would be no reason to include the island.

In the setting of this section, the paradox does not arise for the state of the bulk radiation, but for the boundary state, since we are using the RT prescription to compute the entropy of the latter. A resolution of the state paradox can then be obtained by assuming that the boundary CFT is an *ensemble* of unitary theories, and that the boundary quantities computed using the bulk are ensemble averages (see Fig. 6). This proposal is consistent both with the smallness of $S[R_{d-1}(t)]$ and the fact that the reservoir contains a mixed state, for $t > t_{\mathrm{Page}}$. Since each member of the ensemble is unitary, $S(R_{d-1})$ must follow the Page curve in each theory. Hence the ensemble average of $S(R_{d-1})$ also follows the Page curve.

But the state of R_{d-1} need not be self-averaging. Each member of the ensemble predicts a pure out-state, but this need not be the *same* pure out-state in each theory. Hence the ensemble average of the out-state is a mixed state whose entropy can continue to grow after the Page time. Under the ensemble interpretation, the ensemble-averaged boundary state can be obtained by applying the standard AdS/CFT dictionary to the semiclassical bulk state.

The most explicit calculations of entanglement islands so far [6] were done for the case where the bulk is JT gravity, which is indeed dual to a matrix ensemble. However, we stress that the above argument is unrelated to this observation. The state paradox should be viewed as independent evidence that the gravity path integral, if it is well defined, must be dual to an ensemble, even in settings where no suitable ensemble dual is currently known.

III. GRAVITY/ENSEMBLE DUALITY WITH A BATH

In this section, we turn to the settings studied by Penington [5] and by Almheiri *et al.* [6]. We will argue that the relevant RT prescription can be deduced from the standard one by requiring consistency with the analysis of the previous section. Finally, we will exhibit the state paradox and discuss its resolution by gravity/ensemble duality.

In contrast to Sec. II, the Dyson sphere in AdS_d is eliminated and replaced by an auxiliary (external) system AUX: a "bath" that couples to the boundary CFT_{d-1} and absorbs the Hawking radiation (see Fig. 7). Thus we study the holographic duality

$$M_{d-1} \cup \mathrm{AUX} \to M_d \cup \mathrm{AUX}. \tag{3.1}$$

In Ref. [6], AUX is a $1+1$ dimensional CFT, and the black hole has two asymptotic regions. For definiteness, we will

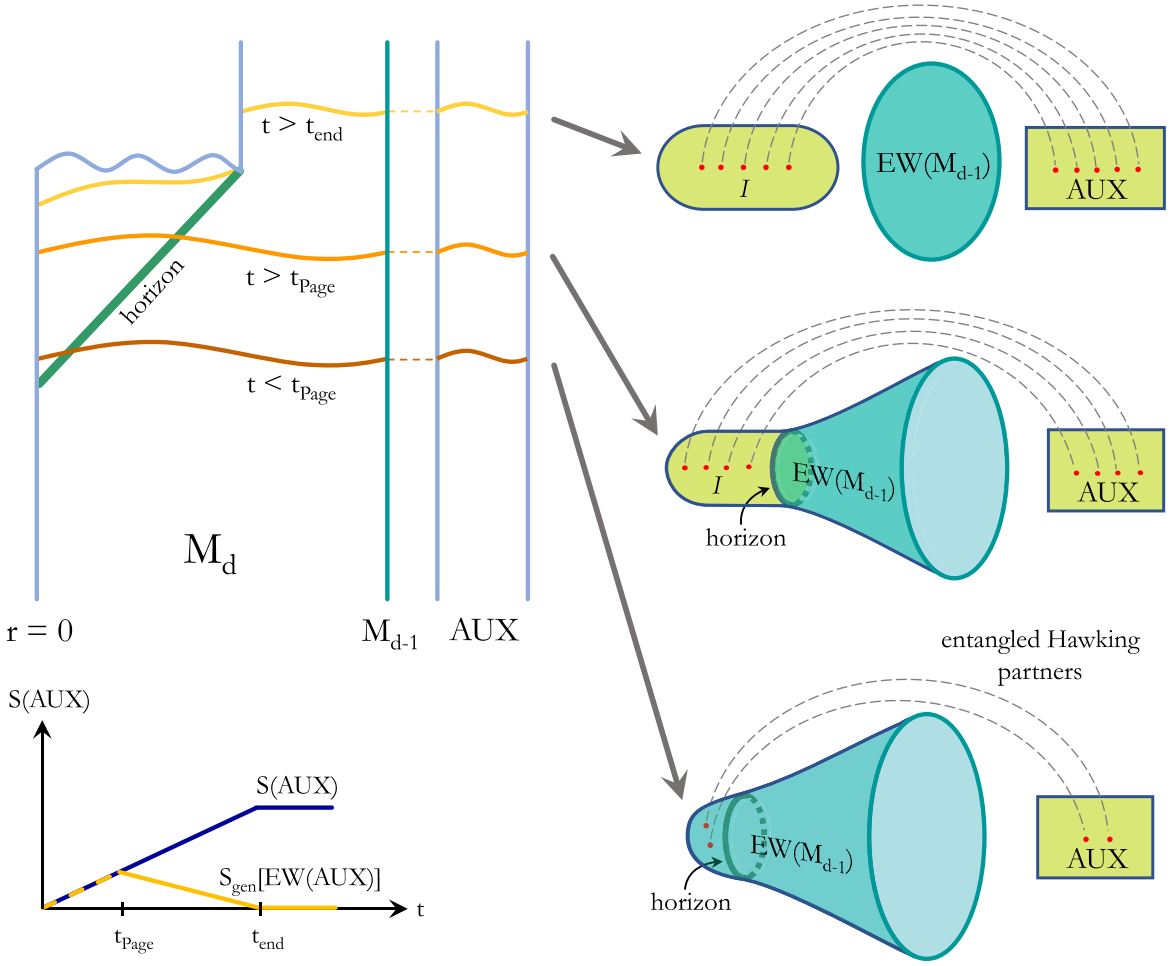

FIG. 7. Hawking radiation escapes into an auxiliary system without gravity. The RT prescription can be applied to the boundary M_{d-1}, yielding the dark green entanglement wedge. A version of the RT prescription for AUX can be developed by requiring consistency with the analysis in Sec. II D. One finds that EW(AUX) (light green) includes AUX itself, and after the Page time, it also the bulk region I complementary to $\mathrm{EW}(M_{d-1})$. The state paradox arises in AUX: the entropy must follow Hawking's rising curve for the island I to be part of EW(AUX), but with I included, RT yields the Page curve for AUX.

follow Penington [5], who considered the more physical setting of a black hole formed from collapse. The auxiliary system AUX will remain unspecified in this section.

We will begin by extending the RT prescription to include AUX, in Sec. III A. Previous efforts to extend the prescription required additional assumptions, such as entanglement wedge complementarity (defined below) [6]. We present a novel argument that this extension is fully determined by the analogy between AUX and the reservoir RES in the previous section. EW complementarity is a consequence rather than an assumption of our argument.

In Sec. III B we apply the RT prescription to black hole evaporation into AUX. This is just for completeness: we summarize Refs. [5,6] and restate the analysis in Sec. II D in this modified setting. The paradox identified in Ref. [22] and reviewed in Sec. II E also has an analogue in this setting. In Sec. III C we discuss this and its resolution if the bulk is dual to an ensemble of boundary theories.

A. Ryu-Takayanagi prescription with auxiliary systems

Consider a bipartite system consisting of a holographic CFT_{d-1} in a region $R_{d-1} \subset M_{d-1}$ and an auxiliary system AUX, in some joint state. Suppose that there exists an RT-like prescription for computing the von Neumann entropy of this state. We shall take "RT-like" to mean that the prescription is of the form

$$S(R_{d-1} \cup \mathrm{AUX}) = S_{\mathrm{gen}}[\mathrm{EW}(R_{d-1} \cup \mathrm{AUX})]. \tag{3.2}$$

We now determine the detailed formulation of the prescription from general considerations.

For any bipartite system consisting of a gravitating region X_d and an auxiliary system AUX, we define the generalized entropy as

$$S_{\mathrm{gen}}(X_d \cup \mathrm{AUX}) = \frac{\mathcal{A}(\partial X_d)}{4G_d} + S(X_d \cup \mathrm{AUX}). \tag{3.3}$$

Given these definitions, the nontrivial content of the prescription we seek lies in how we define the entanglement wedge $\mathrm{EW}(R_{d-1} \cup \mathrm{AUX})$.

Entanglement wedge nesting, the property that the entanglement wedge cannot shrink if the boundary algebra is enlarged [60], implies that

$$\mathrm{EW}(R_{d-1} \cup \mathrm{AUX}) \supset \mathrm{EW}(R_{d-1}). \tag{3.4}$$

We next recall that the relative entropy between two boundary states, $S(\rho|\sigma)$, is the same as the relative entropy between the dual bulk states in the entanglement wedge [61]. This implies that bulk operators in the entanglement wedge (but not outside) can be implemented on the boundary [61–63]. In particular, small deformations of the boundary state do not change the entanglement wedge. Taking $R_{d-1} \cup \mathrm{AUX}$ as the boundary, consider a small deformation of the state in AUX. This can change the boundary relative entropy (in $R_{d-1} \cup \mathrm{AUX}$), but it cannot change the bulk relative entropy in $\mathrm{EW}(R_{d-1} \cup \mathrm{AUX})$, unless we require that

$$\mathrm{EW}(R_{d-1} \cup \mathrm{AUX}) \supset \mathrm{AUX}. \tag{3.5}$$

Therefore, AUX plays an interesting dual role: it appears both on the bulk and on the boundary side.

This does not yet fully determine the prescription. For example, Eqs. (3.4) and (3.5) would be consistent with the (wrong) proposal that $\mathrm{EW}(R_{d-1} \cup \mathrm{AUX})$ is given by $\mathrm{EW}(R_{d-1}) \cup \mathrm{AUX}$. To see that this fails, we note that quantum information can be freely exchanged between AUX and R_{d-1} by appropriate couplings. But consider the setting of Sec. II D. Recall that at the Page time, $\mathrm{EW}(R_{d-1})$ has a phase transition: it now includes not only the portion connected to R_{d-1}, but also an island inside the black hole. Just after the Page time, let us couple the region R_{d-1} to an AUX system that is initially in some pure reference state, and transfer some of the quantum information of the Hawking radiation into AUX. Then the phase transition is reversed; $\mathrm{EW}(R_{d-1})$ loses the island. However, bulk operators in the island could be implemented on $R_{d-1} \cup \mathrm{AUX}$ before the transfer, so this must still be true afterwards. Therefore, $\mathrm{EW}(R_{d-1} \cup \mathrm{AUX})$ cannot have changed.

This shows (at physics-level rigor) that an appropriate definition of the entanglement wedge must treat the bulk and AUX jointly, not separately, when minimizing the generalized entropy. Hence we define

$$\mathrm{EW}(R_{d-1} \cup \mathrm{AUX}) \equiv X_d \cup \mathrm{AUX}, \tag{3.6}$$

where the spacetime region $X_d \subset M_d$ is chosen such that

(1) $\partial X_d = \gamma_d \cup R_{d-1} \cup E_d$, where $\gamma_d \subset M_d - \mathrm{EOW}_d$ and $E_d \subset \mathrm{EOW}_d$.
(2) $S_{\mathrm{gen}}(X_d \cup \mathrm{AUX})$ is stationary under variations of γ_d.
(3) $X_d \cup \mathrm{AUX}$ has the smallest S_{gen} among all regions X_d with the above properties.

We have included the possibility that M_d has an EOW brane for generality, though none appears in the setup studied above. Note that the last term in Eq. (3.3) would vanish in a case where X_d and AUX separately have large von Neumann entropy but purify each other. Note also that AUX in the above formulas could represent one of several auxiliary systems, or equivalently, an arbitrary subalgebra of an auxiliary system.

The generalized RT prescription formulated above upholds entanglement wedge complementarity. Consider a pure quantum state for the complete system $M_{d-1} \cup \mathrm{AUX}$. On the boundary, purity implies $S_{d-1}(R_{d-1} \cup \mathrm{AUX}) = S_{d-1}(\bar{R}_{d-1})$, where $\bar{R}_{d-1}$ is complement of R_{d-1} in M_{d-1}. Purity also implies $\gamma_d(M_{d-1} \cup \mathrm{AUX}) = \varnothing$. Hence $\mathrm{EW}(M_{d-1} \cup \mathrm{AUX}) = M_d \cup \mathrm{AUX}$. The global bulk von Neumann entropy must also vanish: $S(M \cup \mathrm{AUX}) = 0$. This in turn implies that any two subsystems of $M \cup \mathrm{AUX}$ must have equal von Neumann entropy. Therefore $\gamma_d(\bar{R}_{d-1}) = \gamma_d(R_{d-1} \cup \mathrm{AUX})$, and hence

$$\mathrm{EW}(\bar{R}_{d-1}) = \overline{\mathrm{EW}(R_{d-1} \cup \mathrm{AUX})}. \tag{3.7}$$

In the special case where AUX is a nongravitating system described by quantum field theory and $R_{d-1} = \varnothing$, Eq. (3.2) reduces to the "island formula" of Ref. [13], where it was derived using doubly holographic systems. The formula was already used implicitly by Penington [5]. We have argued here that it emerges as a direct consequence of the standard RT prescription, when auxiliary systems are involved. We saw in Sec. II D that the standard RT prescription (without auxiliary systems) already required disconnected islands to be part of the entanglement wedge. The new aspect in the present discussion is not the possibility of an island, but the double role of AUX. This double role will make the state paradox particularly sharp.

B. Island and Page curve

Returning to the specific setting of a black hole evaporating into AUX [5], we now examine the implications of Eq. (3.2). These follow immediately from the results of Sec. II D, upon substituting $R_{d-1} \to \mathrm{AUX}$ and $\bar{R}_{d-1} \to M_{d-1}$. The entropy of each system follows a Page curve, as we will now verify.

Recall that $\Sigma_{d-1}(t)$ defines a foliation of the boundary M_{d-1}, and $\Sigma_d(t)$ are bulk Cauchy slices whose boundary is $\Sigma_{d-1}(t)$. Before the Page time, one finds that the entanglement wedge of the CFT_{d-1} includes the entire bulk:

$$\mathrm{EW}[\Sigma_{d-1}(t)] = \Sigma_d(t). \tag{3.8}$$

Since $\gamma_d = \varnothing$ and hence $\mathcal{A}(\gamma_d) = 0$,

$$S[\Sigma_{d-1}(t)] = S_{\rm gen}[\Sigma_d(t))] = S[\Sigma_d(t)]. \tag{3.9}$$

This grows with time, because $\Sigma_d(t)$ contains the black hole interior, which in turn contains more and more unpartnered interior Hawking modes as the Hawking radiation escapes into AUX.

After the Page time, the entanglement wedge of the full boundary slices $\Sigma_{d-1}(t)$ ends at a quantum extremal surface $\gamma_d(t)$ near the horizon [5]:

$$\mathrm{EW}[\Sigma_{d-1}(t)] = \Sigma_d(t) \cap \mathrm{Ext}[\gamma_d(t)]. \tag{3.10}$$

Here we have chosen $\Sigma_d(t)$ to contain $\gamma_d(t)$, and Ext denotes the spacelike exterior of γ_d. Since the interior Hawking modes are no longer part of $\mathrm{EW}[\Sigma_{d-1}(t)]$, the von Neumann entropy of the entanglement wedge vanishes and so

$$S[\Sigma_{d-1}(t)] = S_{\rm gen}[\mathrm{EW}(\Sigma_{d-1}(t))] = \frac{\mathcal{A}[\gamma_d(t)]}{4G_d}, \tag{3.11}$$

which decreases to zero as the black hole evaporates.

By Eq. (3.7), $\mathrm{EW}[\mathrm{AUX}(t)]$ is the complement of $\mathrm{EW}[\Sigma_{d-1}(t)]$. Thus, the entropy of AUX will follow the same Page curve. Before the Page time, EW(AUX) only contains AUX, i.e., the early Hawking radiation that has been extracted from the AdS_d spacetime. Its entropy grows as more radiation is produced:

$$S_{\rm gen}[\mathrm{EW}(\mathrm{AUX(t)})] = S(\mathrm{AUX}) \qquad (t < t_{\rm Page}). \tag{3.12}$$

After the Page time, EW(AUX) in addition contains an island I:

$$\mathrm{EW}(\mathrm{AUX(t)}) = \mathrm{AUX(t)} \cup I(t), \qquad I = \mathrm{Int}(\gamma_d), \tag{3.13}$$

where Int denotes the spatial interior of γ_d on Σ_d. I is the black hole interior, which contains Hawking partners that purify the radiation in AUX. Hence, the generalized entropy is then given by the (decreasing) boundary area of this island:

$$S_{\rm gen}[\mathrm{EW}(\mathrm{AUX(t)})] = \frac{\mathcal{A}[\gamma_d(t)]}{4G_d} \qquad (t > t_{\rm Page}). \tag{3.14}$$

After the black hole has completely evaporated and all of the Hawking radiation is in AUX, EW(AUX) continues to contain the black hole interior I, now a separate "island universe" without boundary (see Fig. 7).

C. State paradox and ensemble interpretation

In the present setting, the state paradox arises in AUX. On the one hand, the Hawking radiation in AUX is manifestly in a mixed state, whose entropy continues to increase even after the Page time. (In the notation of Ref. [64], this is the "nonbold state.") If its entropy did not increase, then there would be no justification for including the black hole interior island in EW(AUX) after the Page time. On the other hand, the generalized entropy of EW(AUX) after the Page time is given by the area of the black hole, which decreases and eventually vanishes. According to the RT prescription, $S_{\rm gen}[\mathrm{EW(AUX)}]$ computes the von Neumann entropy of AUX. Hence AUX must be in a different state from what we originally assumed: one whose entropy follows the Page curve. (In the notation of Ref. [64], this is the "bold state.") This is a contradiction [22].

It is interesting to compare this instantiation of the state paradox to the version that arose in Sec. II. In Sec. II, the extrapolate dictionary is used at the last step, to translate the mixed Dyson sphere state to a mixed boundary state, in conflict with the pure state obtained from RT. In the present setting, the extrapolate dictionary is used earlier, when the *bulk* Hawking radiation is allowed to escape into AUX by coupling the *boundary* to AUX. Strictly it is not possible to couple radiation inside a spacetime to an auxiliary system, since the resulting nonconservation of the stress tensor would violate the Bianchi identity. Thus the coupling is defined through the boundary, and the extrapolate dictionary is used in interpreting this as a transparent boundary condition for the Hawking radiation. As a result of this coupling, the two conflicting quantum states are both in AUX in the end.

As in the previous section, the paradox is resolved if we assume that the bulk calculation computes both the average state (via Hawking's calculation), and the average entropy (via the RT prescription), in an ensemble of unitary boundary theories. The average entropy of M_{d-1} (and also of AUX) follows the Page curve, because it does so in each (unitary) theory. Different members of the ensemble evolve the same initial state to different final states, so the ensemble average of the state is mixed, and *its* entropy grows monotonically even after the Page time. Both sides of the gravity/ensemble duality exhibit a mixed state: in the bulk because we performed Hawking's calculation, and on the boundary because we averaged over the final state produced by different theories. (In the notation of Ref. [64], the ensemble-averaged bold state equals the nonbold state.)

IV. DOUBLE HOLOGRAPHY WITHOUT A BATH

Beginning with Ref. [13], a number of interesting papers have explored the RT prescription for evaporating black holes in a "doubly holographic" setting [35–39,65]. The Hawking radiation is mainly carried by excitations of a holographic CFT_d that escape to a (holographic) auxiliary system. The state paradox arises in this setting as well, and we will exhibit it in Sec. V. However the analysis is somewhat complicated by the simultaneous appearance of an extra layer of holography and of the auxiliary system.

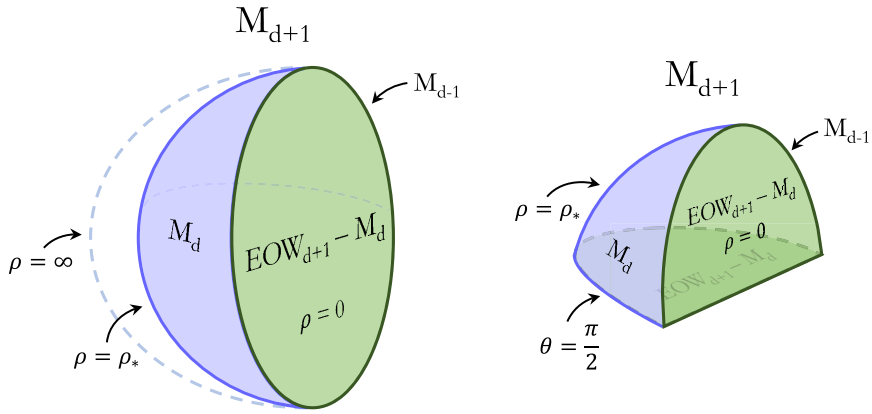

FIG. 8. Double holography without a bath. M_d (purple surface) is the bulk dual of a holographic CFT_{d-1} (left) or BCFT_{d-1} (right) on M_{d-1} (dark green rim). So far this is identical to (a time slice of) the setups shown in Fig. 2. But we now assume that M_d contains a holographic CFT_d. This gives rise to a doubly-holographic bulk dual M_{d+1} (the solid interior). From the $d+1$ bulk perspective, M_d is a Karch-Randall braneworld.

In this section, we will separate these two ingredients: we will introduce double holography without an auxiliary system. We will derive an appropriate "RT-squared" prescription for computing the von Neumann entropy of the top level CFT_{d-1} from its $d+1$ dimensional doubly holographic bulk dual. We will not analyze black hole evaporation and the state paradox in this section; however, our results will be useful when we do so in Sec. V.

A. General setup

As in Sec. II, we consider a holographic CFT_{d-1} on a spacetime M_{d-1}, dual to an asymptotically AdS_d spacetime M_d:

$$M_{d-1} \to M_d. \tag{4.1}$$

We now suppose that the matter sector of the d-dimensional bulk M_d contains a holographic CFT_d coupled to gravity. This implies that M_d is a Randall-Sundrum braneworld [52,66]. The holographic duality can then be iterated:

$$M_d \to M_{d+1}. \tag{4.2}$$

The CFT_d on M_d can be traded for a bulk dual M_{d+1} (see Fig. 8), with Newton's constant G_{d+1} determined by

$$\frac{G_{d+1}}{L_{d+1}} = \frac{G_d}{d-2}. \tag{4.3}$$

Near vacuum regions of the braneworld M_d, M_{d+1} will be locally AdS_{d+1}, with curvature length

$$\frac{L_{d+1}^{d-1}}{G_{d+1}} \sim c_d. \tag{4.4}$$

M_{d+1} will be a manifold with boundary, and we define

$$\mathrm{EOW}_{d+1} = \partial M_{d+1}. \tag{4.5}$$

By definition, the braneworld M_d is a subset of EOW_{d+1}. The complement $\mathrm{EOW}_{d+1} - M_d$ is the boundary of the entanglement wedge of the entire AdS_d brane. Therefore it is located at the minimal-area stationary surface anchored on the AdS_d brane's boundary. It implements boundary conditions on the AdS_{d+1} bulk that are dual the reflecting boundary conditions at the boundary of the AdS_d brane.

The central charge c_d can be thought of as a number of species. In the presence of gravity, large c_d increases the effective Planck length—the cutoff length scale at which the semiclassical analysis breaks down on M_d—from $G_d^{1/(d-2)}$ to $(G_d c_d)^{1/(d-2)} \sim L_{d+1}$. We assume that $G_{d+1}^{1/(d-1)} \ll L_{d+1} \ll L_d$, or equivalently,

$$1 \ll c_d \ll \frac{L_d^{d-2}}{G_d}. \tag{4.6}$$

This ensures that d-dimensional semiclassical gravity is a valid description both in the AdS_{d+1} bulk (the curvature radius is much greater than the Planck scale) and on the AdS_d brane (the curvature radius is much greater than the cutoff scale L_{d+1}).

Usually in holography, there are two descriptions of the same system. The CFT_{d-1} furnishes an exact description. The bulk gives an equivalent description, perturbatively in G_d, in the regime where semiclassical gravity (or perturbative string theory) can be applied. In the setting we consider now, there are three levels:

(1) *Top level:* The CFT_{d-1} on M_{d-1} is the only exact description.
(2) *Holographic bulk dual:* The asymptotically AdS_d bulk M_d with a CFT_d coupled to gravity is an approximate d-dimensional description. Note that

this description is *alternate* to the CFT_{d-1}, so there is no CFT_{d-1} at this level.

(3) *Doubly holographic bulk dual:* The third description, also approximate, is M_{d+1}. There is no CFT_d on the braneworld, at this level; however any other matter fields and dynamical gravity will still be present on M_d.

We will refer to the relation between the top and bottom level as double holography and denote it with a double arrow:

$$M_{d-1} \Rightarrow M_{d+1}. \tag{4.7}$$

Two examples are shown in Fig. 8.

The first example is a holographic CFT_{d-1} on $M_{d-1} = \mathbf{S}^{d-2} \times \mathbf{R}$. In the vacuum state, this is dual to global AdS_d. We now take the AdS_d to contain a holographic CFT_d with the above parameters. Then the CFT_{d-1} has a doubly holographic dual which is locally AdS_{d+1}:

$$\begin{aligned} ds_{d+1}^2 &= L_{d+1}^2[d\rho^2 \\ &\quad + \cosh^2\rho(-\cosh^2 r dt^2 + dr^2 + \sinh^2 r d\Omega_{d-2}^2)], \\ 0 &\le \rho \le \operatorname{arc\,cos\,h}\frac{L_d}{L_{d+1}}. \end{aligned} \tag{4.8}$$

Here $d\Omega_{d-2}^2 = d\theta^2 + \sin^2\theta d\Omega_{d-3}^2$ is the metric on the unit $d-2$ sphere. In these coordinates, the AdS_d brane M_d sits at ρ_* with $\cosh\rho_* = L_d/L_{d+1}$; a second EOW brane resides at $\rho = 0$. See Fig. 8.

The second example of Fig. 8 is half of the previous example. We start with a BCFT_{d-1} on $M_{d-1} = \mathbf{B}^{d-2} \times \mathbf{R}$, where $\mathbf{B}^{d-2}$ is a $d-2$ dimensional hemisphere. For the simplest BCFT with reflecting boundary conditions at the equator, the vacuum state is doubly holographically dual to M_{d+1}, the restriction of Eq. (4.8) to the hemisphere $\theta \le \pi/2$. There is now an additional EOW_{d+1} at $\theta = \pi/2$. The single holographic dual M_d is half of an AdS_d braneworld (still at $\cosh\rho = L_d/L_{d+1}$), with an EOW_d at $\theta = \pi/2$.

B. One-step Ryu-Takayanagi prescription for double holography

The von Neumann entropy S_{d-1} of the CFT_{d-1} restricted to an achronal region $R_{d-1} \subset M_{d-1}$ is given by Eq. (2.5), which we repeat here for convenience:

$$S(R_{d-1}) = S_{\mathrm{gen}}[\mathrm{EW}(R_{d-1})], \tag{4.9}$$

where $\mathrm{EW}(R_{d-1}) \subset M_d$ is the entanglement wedge. In the doubly holographic setting of this section, M_d is a braneworld.

A Ryu-Takayanagi prescription also applies to braneworlds [44,45,67]. Let $R_d \subset M_d$ be an achronal region on the braneworld. Then

$$S_{\mathrm{gen}}(R_d) = S_{\mathrm{gen}}[\mathrm{EW}(R_d)]. \tag{4.10}$$

More generally, R_d may span both a braneworld region and a region (with no gravity) on $\partial\tilde{M}_{d+1}$, the conformal boundary of M_{d+1}; or it may consist of disconnected components in both types of regions. This case will be important in Sec. V; see Fig. 11, with $R_d = \mathrm{EW}(\mathcal{R}_d)$. For $R_d \subset \partial\tilde{M}_{d+1}$, the generalized entropy on the left-hand side is defined as the ordinary von Neumann entropy, with an unregulated UV divergence at ∂R_d. Thus Eq. (4.10) reduces to the usual RT prescription when R_d is entirely on the true boundary.

The entanglement wedge $\mathrm{EW}(R_d)$ is defined as in Sec. II B, with $d \to d+1$: it is an achronal region $X_{d+1} \subset M_{d+1}$, such that

(1) In the unphysical spacetime, $\partial X_{d+1} = \gamma_{d+1} \cup R_d \cup E_{d+1}$. Here $\gamma_{d+1} \subset M_{d+1} - \mathrm{EOW}_{d+1}$, and $E_{d+1} \subset \mathrm{EOW}_{d+1} - R_d$. Note that any portion of R_d that lies on a braneworld is a subset of EOW_{d+1}.

(2) $S_{\mathrm{gen}}(X_{d+1})$ is stationary under variations of γ_{d+1}.

(3) X_{d+1} is has the smallest S_{gen} among all regions with the above properties.

Comparing to Eq. (4.9), an important modification in Eq. (4.10) is that the prescription now computes the *generalized* entropy of the region R_d, rather than purely a CFT_d von Neumann entropy.

The above rules can be combined iteratively, by choosing $R_d = \mathrm{EW}(R_{d-1})$. This allows us to compute any CFT_{d-1} (1st level) entropy using the $d+1$ bulk (the 3rd level). Substituting Eq. (4.10) into Eq. (4.9) we find

$$S(R_{d-1}) = S_{\mathrm{gen}}[\mathrm{EW}(\mathrm{EW}(R_{d-1}))]. \tag{4.11}$$

This is a two-step prescription: one first finds the stationary surface γ_d on the AdS_d brane, and then one finds the stationary surface γ_{d+1} anchored on γ_d. However, we will now show that this is equivalent to simply minimizing the generalized entropy over surfaces that are allowed to be anchored anywhere on the AdS_d brane (and anywhere on the EOW brane), subject to the homology rules described above.

To see this, suppose that the latter procedure yielded a surface γ_{d+1} whose boundary σ on the AdS_d brane was not the minimal QES, γ_d. Then there are two possibilities: (i) σ does not have stationary generalized entropy with respect to small deformations on the brane or (ii) σ is stationary but has larger generalized entropy than γ_d. Case (i) together with the RT rule for braneworlds implies that the generalized entropy of γ_{d+1} (in the $d+1$ bulk) is not stationary under small deformations of γ_{d+1} that reduce to small deformations of σ. Case (ii) implies that the $d+1$ bulk stationary surface anchored on γ_d has smaller generalized entropy than γ_{d+1}. Either of these implications contradicts the definition of γ_{d+1}.

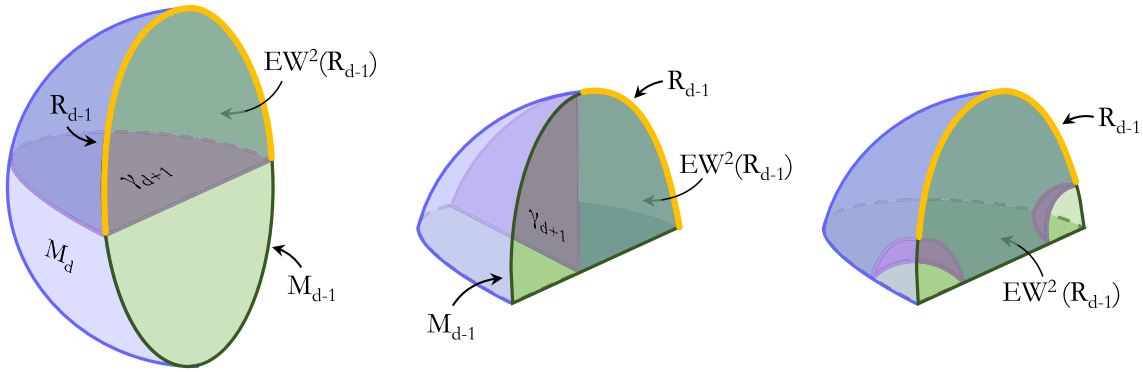

FIG. 9. Examples of the doubly holographic entanglement wedge $\mathrm{EW}^2(R_{d-1})$ for a (B)CFT_{d-1} region R_{d-1}. As before, the light purple surface M_d is the bulk dual of a holographic CFT_{d-1} (left) or BCFT_{d-1} (middle, right) on M_{d-1} (dark green rim). In each case, the doubly holographic entanglement wedge is bounded in part by the surface γ_{d+1}, shown in dark purple.

Thus we can formulate a one-step Ryu-Takayanagi prescription for the von Neumann entropy of a region R_{d-1} of a doubly-holographic CFT_{d-1}:

$$S(R_{d-1}) = S_{\mathrm{gen}}[\mathrm{EW}^2(R_{d-1})], \tag{4.12}$$

where $\mathrm{EW}^2(R_{d-1})$ denotes the *doubly holographic entanglement wedge* of R_{d-1}. This is defined as an achronal region $X_{d+1} \subset M_{d+1}$ such that

(1) In the unphysical spacetime, $\partial X_{d+1} = R_{d-1} \cup \gamma_{d+1} \cup E_{d+1}$. Here $\gamma_{d+1} \subset M_{d+1} - \mathrm{EOW}_{d+1}$ and $E_{d+1} \subset \mathrm{EOW}_{d+1}$.
(2) $S_{\mathrm{gen}}(X_{d+1})$ is stationary under variations of γ_{d+1}.
(3) X_{d+1} has the smallest S_{gen} among all regions with the above properties.

A very simple example is shown in Fig. 9. Consider the CFT_{d-1} in the vacuum state, and let R be half of the $d-1$ sphere in standard global coordinates. Then the QES γ_d is a $d-1$ dimensional hyperbolic plane cutting a Cauchy surface of the AdS_d brane in half: $\cosh\rho = L_d/L_{d+1}; \theta = \pi/2$. (In this example the quantum corrections play no role, so this is also a classical stationary surface.) The QES γ_{d+1} is similarly part of a hyperbolic plane cutting the Cauchy surface of the AdS_{d+1} bulk in half: $\theta = \pi/2$. Of course, it only includes the portion between the AdS_d brane and the EOW brane: $1 < \cosh\rho < L_d/L_{d+1}$. Figure 9 also shows other examples.

C. Quantum vs classical RT in double holography

In the case where the generalized entropies of γ_d and γ_{d+1} are both dominated by the area terms, consistency of Eqs. (4.10) and (4.11) requires

$$\frac{\mathcal{A}(\gamma_d)}{4G_d} = \frac{\mathcal{A}(\gamma_{d+1})}{4G_{d+1}}; \tag{4.13}$$

By Eq. (4.3), this implies a very simple relation between the areas of the QESs:

$$\mathcal{A}(\gamma_{d+1}) = \mathcal{A}(\gamma_d)\frac{2L_{d+1}}{d-2}. \tag{4.14}$$

It is easy to check that this relation is obeyed in the above examples. More generally, consistency requires that γ_{d+1} must have a phase transition if and only if γ_d does, as the region R is varied. For example, if R consists of two antipodal round disks of equal size in the CFT_{d-1}, then γ_d undergoes a well-known phase transition as the disk radius is varied. γ_{d+1} must also have a phase transition at the same critical radius. At first this behavior may seem surprising, because one expects the QESs in the $d+1$ bulk to have a richer structure than those on the AdS_d brane. However, in this context we are only considering $d+1$ QESs anchored on very special surfaces on the AdS_d brane—those that are themselves QESs—so there is no contradiction.

A more interesting case arises when the CFT_d is far from its vacuum state, so that the von Neumann entropy of braneworld regions is large. In this case S_{gen} on M_d may have large quantum contributions (i.e., contributions from the von Neumann entropy term), while S_{gen} of the corresponding entanglement wedge in M_{d+1} is dominated by the classical term (the area term). In such a case, one can replace S_{gen} by $\mathcal{A}(\gamma_{d+1})/4G_{d+1}$ in Eqs. (4.10)–(4.12), but not by $\mathcal{A}(\gamma_d)/4G_d$ in Eq. (4.9).

V. DOUBLE HOLOGRAPHY WITH A HOLOGRAPHIC BATH

This section can be thought of as an extension of the above settings, in two different ways. Continuing from the previous section, we keep the doubly holographic setup but we add a bath. That is, we couple the CFT_{d-1} (or equivalently, the AdS_d brane) to an auxiliary system AUX. We take AUX to be the same holographic CFT_d that lives on the AdS_d brane, but not coupled to gravity. Thus AUX can be thought of as a CFT_d living on a true asymptotic boundary of an asymptotically AdS_{d+1} bulk dual.

From the perspective of Sec. III, we keep the bath but make the setting doubly holographic. That is, we now

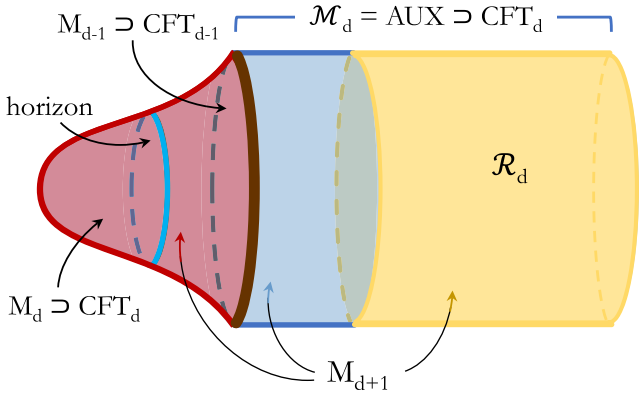

FIG. 10. A doubly holographic CFT_{d-1} on M_{d-1} is coupled to holographic bath: a CFT_d on $\mathcal{M}_d$. The first holographic dual is $M_d \cup \mathcal{M}_d$, where M_d contains the same CFT_d coupled to gravity. The second holographic dual is M_{d+1} (solid interior). We consider a state which, in the first dual, corresponds to an evaporating black hole in M_d with radiation escaping to $\mathcal{M}_d$. The von Neumann entropy of the radiation in the subregion $\mathcal{R}_d \subset \mathcal{M}_d$ can be computed using the single or double RT prescription.

specialize to the case where both the dominant matter content in the gravitating AdS_d spacetime, and also the external bath AUX is a holographic CFT_d, with an asymptotically AdS_{d+1} bulk dual.

Combining insights from the previous sections will allow us to understand some puzzling features in the doubly-holographic versions [13,35–39] of Refs. [5,6], where the Page curve arises from the classical RT prescription in the $d+1$ dimensional bulk. One such feature is the appearance of two apparently different states in the bath region, denoted bold and nonbold in Ref. [13]. We will see that these states need not be different in the ensemble interpretation.

A. General setup

As before, we consider a holographic CFT_{d-1} with central charge c_{d-1} on a manifold M_{d-1}, dual to a bulk M_d. We now choose this CFT_{d-1} such that the matter content of M_d includes a particular CFT_d coupled to gravity. As in Sec. III, we couple the CFT_{d-1} to an auxiliary system AUX (see Fig. 10). We now insist that AUX is specifically a CFT_d on a manifold $\mathcal{M}_d$ such that $M_{d-1} = \partial\mathcal{M}_d$, and we take this to be the same CFT_d that also appears in the bulk dual M_d.

The coupled boundary system (CFT_{d-1} on M_{d-1} and CFT_d on $\mathcal{M}_d$) defines a BCFT_d on $\mathcal{M}_d$. Importantly, there is no dynamical gravity on $\mathcal{M}_d$. Applying the general discussion of Sec. III to the CFT_{d-1} and AUX (i.e., to the BCFT_d), we find that this system is holographically dual to a d-dimensional bulk system:

$$M_{d-1} \cup \mathcal{M}_d \to M_d \cup \mathcal{M}_d. \tag{5.1}$$

Here M_d has dynamical gravity. AUX $= \mathcal{M}_d$ plays a dual role as bulk and boundary system.

Next, we add the ingredient of double holography, as in Sec. IV. Suppose that the CFT_d on $M_d \cup \mathcal{M}_d$ is holographic, with parameters as described in Sec. IV A. Let M_{d+1} be its $d+1$ dimensional bulk dual:

$$M_d \cup \mathcal{M}_d \to M_{d+1}. \tag{5.2}$$

As usual, let $\tilde{M}_{d+1}$ be the associated unphysical spacetime (Penrose diagram), and let $\mathrm{EOW}_{d+1} = \partial M_{d+1}$. Then $\mathcal{M}_d = \partial\tilde{M}_{d+1}$ and $M_d \subset \mathrm{EOW}_{d+1}$. The above two dualities combine to establish the doubly holographic duality

$$M_{d-1} \cup \mathcal{M}_d \Rightarrow M_{d+1}. \tag{5.3}$$

For example, with $M_{d-1} = \mathbf{S}^{d-2} \times \mathbf{R}$ at the equator of the hemishere $\mathcal{M}_d = \mathbf{B}^{d-1} \times \mathbf{R}$, one obtains the Karch-Randall (KR) model [66]. This was first discussed in detail as a doubly-holographic model in Ref. [68]. The first bulk dual is $M_d \cup \mathcal{M}_d$, where M_d is an AdS_d braneworld known as a KR brane. It forms the boundary of the doubly holographic dual M_{d+1}, a global AdS_{d+1} spacetime that terminates on the KR brane. In the vacuum state, the metric of M_{d+1} is given by Eq. (4.8), with the range of ρ extended to

$$-\infty < \rho \leq \operatorname{arccosh}\frac{L_d}{L_{d+1}}; \tag{5.4}$$

The braneworld M_d is located at the upper end of this range, and the asymptotic boundary $\mathcal{M}_d$ is at the lower end. M_{d-1} is at $\rho = 0$, $r \to \infty$.

Alternatively, let $M_{d-1} = \mathbf{R}^{d-2} \times \mathbf{R}$ be the boundary of the half-space $\mathcal{M}_d = \mathbf{B}^{d-1} \times \mathbf{R}$. This gives the Poincaré patch of an AdS_d braneworld as the first bulk dual, M_d; it gives the Poincaré patch of AdS_{d+1} as the second bulk dual M_{d+1}.

Both of these models were studied further by Takayanagi and collaborators [42,43], who gave a one-step RT prescription for the duality in Eq. (5.3). We will now derive this prescription from a different perspective, by combining the results of the previous sections.

B. One-step Ryu-Takayanagi prescription for double holography

The one-step RT prescription for the doubly holographic duality (5.3) can be derived iteratively by combining the RT prescriptions for the single holographic dualities (5.1) and (5.2). For the first step this was given in Eqs. (3.2)–(3.6). Setting AUX $\to \mathcal{R}_d$, Eq. (3.2) becomes:

$$S(R_{d-1} \cup \mathcal{R}_d) = S_{\mathrm{gen}}[\mathrm{EW}(R_{d-1} \cup \mathcal{R}_d)], \tag{5.5}$$

where $R_{d-1} \subset M_{d-1}$ and $\mathcal{R}_d \subset \mathcal{M}_d - M_{d-1}$ are arbitrary subregions of the boundary system. The other equations

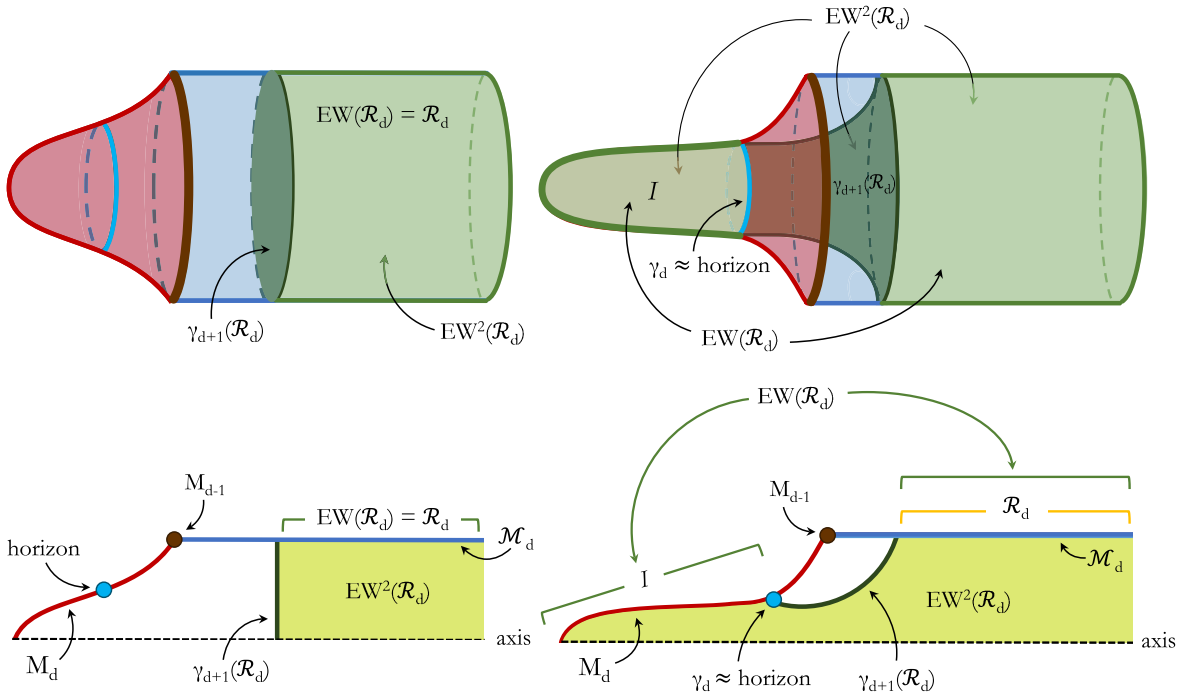

FIG. 11. Entanglement wedges of the bath region $\mathcal{R}_d$ before (left) and after (right) the Page time, when $\mathrm{EW}(\mathcal{R}_d)$ has a disconnected island I. Each top figure is simply the bottom figure rotated around the axis. The "squared" entanglement wedge EW^2 is always connected. It can be found iteratively as $\mathrm{EW}(\mathrm{EW}(\mathcal{R}_d))$, or in one step from Eq. (5.8) [42,43]. As in Sec. V C, γ_d is a quantum extremal surface, but γ_{d+1} is an ordinary extremal surface.

and the definition of EW are as in Sec. III. The fact that the auxiliary system is a field theory plays no role in this step.

The second step computes the generalized entropy on the RHS of Eq. (5.5) holographically. Setting $R_d \to \mathrm{EW}(R_{d-1} \cup \mathcal{R}_d)$ in Eq. (4.10), we obtain

$$S_{\mathrm{gen}}(\mathrm{EW}(R_{d-1} \cup \mathcal{R}_d)) = S_{\mathrm{gen}}[\mathrm{EW}(\mathrm{EW}(R_{d-1} \cup \mathcal{R}_d))]. \tag{5.6}$$

Thus we obtain

$$S(R_{d-1} \cup \mathcal{R}_d) = S_{\mathrm{gen}}[\mathrm{EW}(\mathrm{EW}(R_{d-1} \cup \mathcal{R}_d))]. \tag{5.7}$$

By arguments exactly analogous to those following Eq. (4.11), this iterative result can be condensed into a one-step RT prescription:

$$S(R_{d-1} \cup \mathcal{R}_d) = S_{\mathrm{gen}}[\mathrm{EW}^2(R_{d-1} \cup \mathcal{R}_d)]. \tag{5.8}$$

The doubly-holographic entanglement wedge $\mathrm{EW}^2(R_{d-1} \cup \mathcal{R}_d)$ is defined as an achronal region $X_{d+1} \subset M_{d+1}$ such that

(1) In the unphysical spacetime, $\partial X_{d+1} = R_{d-1} \cup \mathcal{R}_d \cup \gamma_{d+1} \cup E_{d+1}$, where $\gamma_{d+1} \subset M_{d+1} - \mathrm{EOW}_{d+1}$ and $E_{d+1} \subset \mathrm{EOW}_{d+1}$.
(2) $S_{\mathrm{gen}}(X_{d+1})$ is stationary under variations of γ_{d+1}.
(3) X_{d+1} is has the smallest S_{gen} among all regions with the above properties.

We note that this agrees with the RT prescription for a BCFT_d given by Takayanagi [42,43], which has been extensively used in recent analyses of entanglement islands, such as Refs. [13,35–39]. In analyzing these results and exhibiting the state paradox, it will be illuminating to "deconstruct" Eq. (5.8) in to Eqs. (5.5) and (5.6).

C. Island and Page curve

We now specialize to the dynamical setting of Sec. III B: the first holographic dual, $M_d \cup \mathcal{M}_d$, M_d contains a black hole whose radiation propagates to $\mathcal{M}_d$. First, let us consider the top level, the BCFT_d on $M_{d-1} \cup \mathcal{M}_d$. There are now two ways to compute the von Neumann entropy of a subregion $\mathcal{R}_d \subset \mathcal{M}_d$ that contains the radiation.

One option is to ignore the second holographic dual and use RT only for the first holographic duality, Eq. (5.1). Setting $R_{d-1} \to \varnothing$ in Eq. (5.5), we find

$$S(\mathcal{R}_d) = S_{\mathrm{gen}}[\mathrm{EW}(\mathcal{R}_d)]. \tag{5.9}$$

Before the Page time, $\mathrm{EW}(\mathcal{R}_d) = \mathcal{R}_d$ (see Fig. 11). Since $\mathcal{R}_d$ is a true boundary region, $S_{\mathrm{gen}}(\mathcal{R}_d) = S(\mathcal{R}_d)$. Thus, the above equation is a trivial identity before the Page time. After the Page time,

$$\mathrm{EW}(\mathcal{R}_d) = \mathcal{R}_d \cup I \qquad (t > t_{\mathrm{Page}}), \tag{5.10}$$

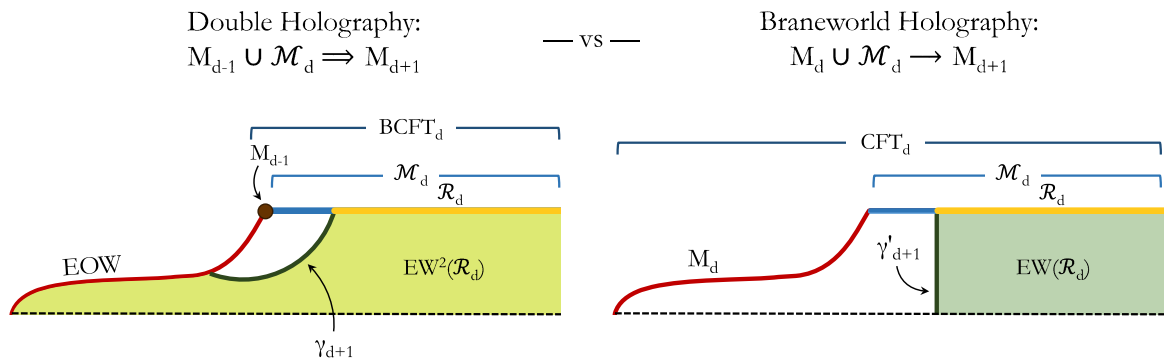

FIG. 12. Bulk dual of the state paradox. Left: we regard $\mathcal{R}_d$ as a BCFT_d subregion (top level). The homology rule following Eq. (5.8) applies: γ_{d+1} is allowed to end on the braneworld which here appears as an EOW brane. At late times, $\mathcal{A}(\gamma_{d+1}) \to 0$, resulting in the Page curve for $S(\mathcal{R}_d)$. Right: we consider $\mathcal{R}_d$ as a subregion of the CFT_d on $M_d \cup \mathcal{M}_d$. The homology rule following Eq. (4.10) applies. The braneworld M_d is now part of the boundary; since we are computing the entropy only for the region $\mathcal{R}_d$, γ'_{d+1} is not allowed to end on M_d. $\mathcal{A}(\gamma'_{d+1})$ grows monotonically, resulting in Hawking's curve.

where the island $I \subset M_d$ is the black hole interior (see Fig. 11). The Hawking radiation in $\mathcal{R}_d$ is purified by the Hawking partners in I, so

$$S_{\mathrm{gen}}[\mathrm{EW}(\mathcal{R}_d)] = \frac{\mathcal{A}(\gamma_d)}{4G_d} \qquad (t > t_{\mathrm{Page}}), \tag{5.11}$$

where $\gamma_d = \partial I$ nearly coincides with the horizon. Note that the radiation appears on both sides of the duality, and that we have made no reference to the second holographic bulk dual M_{d+1}.

Another option is to use the doubly holographic duality, Eq. (5.3). By Eq. (5.8),

$$S(\mathcal{R}_d) = S_{\mathrm{gen}}[\mathrm{EW}^2(\mathcal{R}_d)]. \tag{5.12}$$

With the one-step prescription following Eq. (5.8) one finds $\mathrm{EW}^2(\mathcal{R}_d)$ as shown in Fig. 11. Unlike $\mathrm{EW}(\mathcal{R}_d)$ in Eq. (5.10), $\mathrm{EW}^2(\mathcal{R}_d)$ is always a connected region. After the Page time, γ_{d+1} ends on the quantum extremal surface γ_d, and the island I forms part of the boundary of $\mathrm{EW}^2(\mathcal{R}_d)$. But neither the radiation in $\mathcal{R}_d$ nor the Hawking partners in the black hole interior on M_d contribute to $S_{\mathrm{gen}}[\mathrm{EW}^2(\mathcal{R}_d)]$, since they are not part of M_{d+1}. Both before and after the Page time, the generalized entropy of the squared entanglement wedge is given just by the classical area of γ_{d+1}, in line with the discussion at the end of Sec. IV C:

$$S_{\mathrm{gen}}[\mathrm{EW}^2(\mathcal{R}_d(t))] = \frac{\mathcal{A}[\gamma_{d+1}(t)]}{4G_{d+1}}. \tag{5.13}$$

D. State paradox and ensemble interpretation

Agreement between Eqs. (5.9) and (5.12) is a nontrivial consequence of Eq. (5.6). That equation, in turn, was obtained by applying the RT prescription for braneworlds, (4.10), which is relevant for the duality (5.2), to the region $\mathrm{EW}(\mathcal{R}_d)$. But Eq. (4.10) allows us to choose any other subregion of the first bulk dual $M_d \cup \mathcal{M}_d$ and compute its generalized entropy. Thus we may ask questions that have no obvious analogue in the dualities of Eqs. (5.1) and (5.3).

For example, after the Page time, $\mathrm{EW}(\mathcal{R}_d) = \mathcal{R}_d \cup I$. But we could instead use Eq. (4.10) to compute the generalized entropy of just $\mathcal{R}_d$. Because Eq. (4.10) prohibits the RT surface γ'_{d+1} from ending on M_d (see Fig. 12), its area continues to grow after the Page time, and we find the entropy computed by Hawking. Thus, Eq. (4.10) will not give the same answer for $S(\mathcal{R}_d)$ as Eqs. (5.9) and (5.12). This contradiction is the bulk dual of the state paradox.

In Sec. III C (with $\mathrm{AUX} \to \mathcal{R}_d$), the state paradox appeared as a contradiction between $S(\mathcal{R}_d)$ computed from the semiclassical Hawking analysis on $M_d \cup \mathcal{M}_d$, and $S(\mathcal{R}_d)$ computed from Eq. (5.9). Either quantity can now also be computed using the second holographic dual M_{d+1}. As noted in the previous paragraph, the results (given by Eq. (4.10) and (5.12) respectively) disagree.

Gravity/ensemble duality can again resolve this paradox. Suppose that the CFT_{d-1} on M_{d-1} is really an ensemble of unitary theories as discussed in the introduction. From the top-level viewpoint, the CFT_{d-1} emits radiation into the CFT_d on $\mathcal{M}_d$. In each theory, this process is unitary and the radiation entropy in $\mathcal{R}_d \subset \mathcal{M}_d$ follows the Page curve. Hence the average entropy $\langle S(\rho_{\mathcal{R}_d})\rangle$ follows the Page curve. But the ensemble-averaged state of the radiation, $S(\langle \rho_{\mathcal{R}_d}\rangle)$, follows Hawking's monotonically rising curve.

The first holographic dual of this process is the escape of Hawking radiation from M_d into $\mathcal{M}_d$. Assuming gravity/ensemble duality, the semiclassical analysis of black hole evaporation computes $\langle \rho_{\mathcal{R}_d}\rangle$ directly, and it determines $\langle S(\rho_{\mathcal{R}_d})\rangle$ via the first RT prescription, Eq. (5.9). The second layer of holography, Eq. (5.2), gives us an alternative way of computing $\langle S(\rho_{\mathcal{R}_d})\rangle$ and $S(\langle \rho_{\mathcal{R}_d}\rangle)$ using the braneworld

version of the RT prescription, Eq. (4.10). To compute $\langle S(\rho_{\mathcal{R}_d})\rangle$, choose $R_d \to \mathrm{EW}(\mathcal{R}_d) = \mathcal{R}_d \cup I$ in Eq. (4.10). To compute $S(\langle\rho_{\mathcal{R}_d}\rangle)$, set $R_d \to \mathcal{R}_d$ in Eq. (4.10).

It is interesting to note that it does not matter whether Eq. (5.2) is a gravity/ensemble duality. Suppose that it is. Then there exists an ensemble of CFT_d theories on $M_d \cup \mathcal{M}_d$. On what is now the boundary side, we would have to perform a gravity path integral involving each of these different theories, then average. But regardless of the details of each CFT_d, the state in $\mathcal{R}_d$ will be thermal and purified by the excitation in I. Therefore, unlike the state of the Hawking radiation in $\mathcal{R}_d$ in the BCFT_d (the top level), the state of the semiclassically evolved CFT_d theories is self-averaging in the region $\mathcal{R}_d \cup I$, and also in the region $\mathcal{R}_d$. Of course, in a different state (for example, a setup analogous to Sec. II in the $d+1$ dimensional bulk), a state paradox can arise in $M_d \cup \mathcal{M}_d$, and we would need to appeal to state gravity/ensemble duality for a resolution.

So far, we have discussed the first and second holographic duality separately. We can also consider the one-step doubly holographic RT prescription of Eq. (5.8). This evaluates the entropy of the BCFT_d region $\mathcal{R}_d$ directly in the M_{d+1} bulk as the area of γ_{d+1}; see Eq. (5.13). By "jumping" over the middle level, we have missed the paradox. Namely, the paradox involved the apparent discrepancy of the states in the region $\mathcal{R}_d$, depending on whether it is viewed as a state of the BCFT_d or a state of the CFT_d on $M_d \cup \mathcal{M}_d$. The CFT_d is not present in the second holographic dual. It has now been replaced by the classical bulk state in M_{d+1}; thus we are no longer comparing two states of the same region.

Therefore, an ensemble interpretation is not required to make sense of the doubly holographic duality (5.3), so long as we never consider the intermediate level. Unfortunately, without the intermediate level $M_d \cup \mathcal{M}_d$, we also lose contact with the process of black hole evaporation, which is manifest only at this level.

ACKNOWLEDGMENTS

We would like to thank C. Akers, A. Almheiri, N. Bao, V. Chandrasekaran, R. Emparan, N. Engelhardt, T. Hartman, A. Levine, R. Mahajan, J. Maldacena, A. Maloney, H. Marrochio, H. Maxfield, M. Miyaji, G. Penington, A. Shahbazi-Moghaddam, R. Myers, P. Rath, G. Remmen, J. Santos, D. Stanford, M. Tomašević and Y. Zhao for helpful discussions, comments, and correspondence. This work was supported in part by the Berkeley Center for Theoretical Physics; by the Department of Energy, Office of Science, Office of High Energy Physics under QuantISED Award No. DE-SC0019380 and under contract No. DE-AC02-05CH11231; and by the National Science Foundation under Grant No. PHY1820912.

[1] J. Polchinski, *The Black Hole Information Problem* (World Scientific, Singapore, 2017), Chap. 6, pp. 353–397.
[2] S. W. Hawking, Black hole explosions, Nature (London) **248**, 30 (1974).
[3] S. W. Hawking, Black holes and thermodynamics, Phys. Rev. D **13**, 191 (1976).
[4] J. Maldacena, The large N limit of superconformal field theories and supergravity, Adv. Theor. Math. Phys. **2**, 231 (1998).
[5] G. Penington, Entanglement wedge reconstruction and the information paradox, arXiv:1905.08255.
[6] A. Almheiri, N. Engelhardt, D. Marolf, and H. Maxfield, The entropy of bulk quantum fields and the entanglement wedge of an evaporating black hole, J. High Energy Phys. 12 (2019) 063.
[7] A. Lewkowycz and J. Maldacena, Generalized gravitational entropy, J. High Energy Phys. 08 (2013) 090.
[8] S. Ryu and T. Takayanagi, Holographic Derivation of Entanglement Entropy from AdS/CFT, Phys. Rev. Lett. **96**, 181602 (2006).
[9] S. Ryu and T. Takayanagi, Aspects of holographic entanglement entropy, J. High Energy Phys. 08 (2006) 045.
[10] V. E. Hubeny, M. Rangamani, and T. Takayanagi, A covariant holographic entanglement entropy proposal, J. High Energy Phys. 07 (2007) 062.
[11] T. Faulkner, A. Lewkowycz, and J. Maldacena, Quantum corrections to holographic entanglement entropy, J. High Energy Phys. 11 (2013) 074.
[12] N. Engelhardt and A. C. Wall, Extremal surface barriers, J. High Energy Phys. 03 (2014) 068.
[13] A. Almheiri, R. Mahajan, J. Maldacena, and Y. Zhao, The page curve of Hawking radiation from semiclassical geometry, J. High Energy Phys. 03 (2020) 149.
[14] A. Almheiri, D. Marolf, J. Polchinski, and J. Sully, Black holes: Complementarity or firewalls?, J. High Energy Phys. 02 (2013) 062.
[15] G. Penington, S. H. Shenker, D. Stanford, and Z. Yang, Replica wormholes and the black hole interior, arXiv:1911.11977.
[16] A. Almheiri, T. Hartman, J. Maldacena, E. Shaghoulian, and A. Tajdini, Replica wormholes and the entropy of Hawking radiation, J. High Energy Phys. 05 (2020) 013.
[17] A. Almheiri, T. Hartman, J. Maldacena, E. Shaghoulian, and A. Tajdini, The entropy of Hawking radiation, arXiv:2006.06872.

[18] T. Hartman, E. Shaghoulian, and A. Strominger, Islands in asymptotically flat 2D gravity, J. High Energy Phys. 07 (2020) 022.

[19] F. F. Gautason, L. Schneiderbauer, W. Sybesma, and L. Thorlacius, Page curve for an evaporating black hole, J. High Energy Phys. 05 (2020) 091.

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

[21] C. Akers, N. Engelhardt, and D. Harlow, Simple holographic models of black hole evaporation, J. High Energy Phys. 08 (2020) 032.

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

[23] C. Teitelboim, Gravitation and Hamiltonian structure in two space-time dimensions, Phys. Lett. B **126**, 41 (1983).

[24] R. Jackiw, Lower dimensional gravity, Nucl. Phys. **B252**, 343 (1985).

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

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

[27] D. Stanford and E. Witten, JT gravity and the ensembles of random matrix theory, arXiv:1907.03363.

[28] P. Saad, Late time correlation functions, baby universes, and ETH in JT gravity, arXiv:1910.10311.

[29] E. Witten, Matrix models and deformations of JT gravity, arXiv:2006.13414.

[30] N. Afkhami-Jeddi, H. Cohn, T. Hartman, and A. Tajdini, Free partition functions and an averaged holographic duality, arXiv:2006.04839.

[31] A. Maloney and E. Witten, Averaging over Narain moduli space, arXiv:2006.04855.

[32] J. Cotler and K. Jensen, AdS_3 gravity and random CFT, arXiv:2006.08648.

[33] A. Prez and R. Troncoso, Gravitational dual of averaged free CFT's over the Narain lattice, arXiv:2006.08216.

[34] H. Maxfield and G. J. Turiaci, The path integral of 3D gravity near extremality; or, JT gravity with defects as a matrix integral, arXiv:2006.11317.

[35] M. Rozali, J. Sully, M. Van Raamsdonk, C. Waddell, and D. Wakeham, Information radiation in BCFT models of black holes, J. High Energy Phys. 05 (2020) 004.

[36] H. Z. Chen, Z. Fisher, J. Hernandez, R. C. Myers, and S.-M. Ruan, Information flow in black hole evaporation, J. High Energy Phys. 03 (2020) 152.

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

[38] J. Sully, M. Van Raamsdonk, and D. Wakeham, BCFT entanglement entropy at large central charge and the black hole interior, arXiv:2004.13088.

[39] D. Bak, C. Kim, S.-H. Yi, and J. Yoon, Unitarity of entanglement and islands in two-sided Janus black holes, arXiv:2006.11717.

[40] J. Cardy, Boundary conformal field theory, in *Encyclopedia of Mathematical Physics*, edited by J.-P. Franoise, G. L. Naber, and T. S. Tsun (Academic Press, Oxford, 2006), pp. 333–340.

[41] I. Affleck and A. W. W. Ludwig, Universal Noninteger 'Ground State Degeneracy' in Critical Quantum Systems, Phys. Rev. Lett. **67**, 161 (1991).

[42] T. Takayanagi, Holographic Dual of BCFT, Phys. Rev. Lett. **107**, 101602 (2011),

[43] M. Fujita, T. Takayanagi, and E. Tonni, Aspects of AdS/BCFT, J. High Energy Phys. 11 (2011) 043,

[44] R. Emparan, Black hole entropy as entanglement entropy: A holographic derivation, J. High Energy Phys. 06 (2006) 012.

[45] R. C. Myers, R. Pourhasan, and M. Smolkin, On spacetime entanglement, J. High Energy Phys. 06 (2013) 013.

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

[47] A. Almheiri, D. Marolf, J. Polchinski, D. Stanford, and J. Sully, An apologia for firewalls, J. High Energy Phys. 09 (2013) 018.

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

[49] R. Bousso, Complementarity is not enough, Phys. Rev. D **87**, 124023 (2013).

[50] R. Bousso, Firewalls from double purity, Phys. Rev. D **88**, 084035 (2013).

[51] R. Bousso, Violations of the Equivalence Principle by a Nonlocally Reconstructed Vacuum at the Black Hole Horizon, Phys. Rev. Lett. **112**, 041102 (2014).

[52] L. Randall and R. Sundrum, A Large Mass Hierarchy from a Small Extra Dimension, Phys. Rev. Lett. **83**, 3370 (1999).

[53] R. M. Wald, *General Relativity* (University of Chicago Press, Chicago, 1984).

[54] I. Kourkoulou and J. Maldacena, Pure states in the SYK model and nearly-AdS_2 gravity, arXiv:1707.02325.

[55] S. Cooper, M. Rozali, B. Swingle, M. Van Raamsdonk, C. Waddell, and D. Wakeham, Black hole microstate cosmology, J. High Energy Phys. 07 (2019) 065.

[56] N. Engelhardt and A. C. Wall, Quantum extremal surfaces: Holographic entanglement entropy beyond the classical regime, J. High Energy Phys. 01 (2015) 073.

[57] J. D. Bekenstein, Black holes and the second law, Nuovo Cimento Lett. **4**, 737 (1972).

[58] D. A. Lowe, J. Polchinski, L. Susskind, L. Thorlacius, and J. Uglum, Black hole complementarity versus locality, Phys. Rev. D **52**, 6997 (1995).

[59] A. Hamilton, D. N. Kabat, G. Lifschytz, and D. A. Lowe, Holographic representation of local bulk operators, Phys. Rev. D **74**, 066009 (2006).

[60] A. C. Wall, Maximin surfaces, and the strong subadditivity of the covariant holographic entanglement entropy, Classical Quantum Gravity **31**, 225007 (2014).

[61] D. L. Jafferis, A. Lewkowycz, J. Maldacena, and S. J. Suh, Relative entropy equals bulk relative entropy, J. High Energy Phys. 06 (2016) 004.

[62] A. Almheiri, X. Dong, and D. Harlow, Bulk locality and quantum error correction in AdS/CFT, J. High Energy Phys. 04 (2015) 163.

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

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

[65] 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, arXiv:2006.04851.

[66] A. Karch and L. Randall, Locally localized gravity, J. High Energy Phys. 05 (2001) 008.

[67] J. Koeller and S. Leichenauer, Holographic proof of the quantum null energy condition, Phys. Rev. D **94**, 024026 (2016).

[68] R. Bousso and L. Randall, Holographic domains of Anti-de Sitter space, J. High Energy Phys. 04 (2002) 057.