# The entanglement wedge duality and Hilbert space factorization in AdS/CFT, Karch-Randall braneworld and black hole physics

Minseong Kim*

(Dated: December 31, 2024)

## Abstract

The entanglement wedge duality in AdS/CFT, Karch-Randall braneworld and black hole physics is discussed, critically (as in critiques) revolving around Terashima (2024) and Mathur (2009). Emphasis are placed on the semiclassicality requirement and potential breakdown of bulk-wise effective Hilbert space factorization between the black hole exterior and the interior, supported by various examples. In particular, a simple quasi-paradox for the entanglement wedge duality in AdS/BCFT is put forward, suggesting that we cannot have all of semiclassicality, the entanglement wedge duality and bulk-wise Hilbert space factorization in Karch-Randall braneworld. Giving up factorization, a toy qudit model of BCFT can be developed as to derive purification of the black hole exterior, despite mostly sharing the setup with the small corrections theorem of Mathur (2009). Put together, the entanglement wedge duality is yet to be disproved despite some results by Seiji Terashima, and a toy qudit model cannot disprove semiclassicality in black hole physics. The precise understanding of the duality also helps identify how the black hole information paradox is dissolved within double holography beyond Karch-Randall braneworld setups.

Keywords: entanglement wedge duality, Hilbert space factorization, Karch-Randall braneworld, semiclassical physics, black hole information paradox, small corrections theorem, toy qubit model

* mkimacad@gmail.com; ORCiD:0000-0003-2115-081X

**CONTENTS**



## I. INTRODUCTION

In this paper, validity of the entanglement wedge duality [1] and the small corrections theorem [2] are discussed.

The entanglement wedge duality in AdS/CFT has recently been heavily criticized by Seiji Terashima [3–6], and we discuss how the duality remains yet to be disproved when restricted to semiclassical physics. In Karch-Randall braneworld, in case massless gravity is assumed for brane $Q$, the entanglement wedge duality has been criticized as being invalid [7]. These backgrounds provide an impetus for formulating an information quasi-paradox in Section III A, which provides a simpler Hilbert space-wise critique of the entanglement wedge duality in Karch-Randall braneworld unless Hilbert space factorization of the bulk is quite different from that of the boundary.

Nevertheless, we will see that as long as we accept non-trivial Hilbert space factorization differences between the bulk and the dual boundary theory, all of these critiques are addressed. Indeed, Terashima (2024) [6] provides an example where such differences can clearly be noted, which we discuss in Section II. These differences show why the small corrections theorem [2] becomes conceptually inapplicable in holography with 'boundary of boundary' (double holography).

Assuming breakdown of bulk-wise Hilbert space factorization between the black hole exterior and the interior, we further develop a toy BCFT qudit model in Section III D, similar to the setup of the small corrections theorem [2], that explicitly shows how the black hole exterior is purified. This can be considered as simplification of the moving mirrors model [8, 9] but also a generalization in sense that we do not require specific Karch-Randall braneworld setups. The toy model suggests why a toy qudit model cannot be used to disprove semiclassicality in black hole physics, especially within holography with 'boundary of boundary' (double holography). Alternatively, we could understand the small corrections theorem as demonstrating that bulk-wise Hilbert space factorization between the black hole interior and the exterior has to be sacrificed if we are to preserve semiclassicality in black hole physics.

The lessons to be obtained at the end go as follows:

- The entanglement wedge duality remains valid for semiclassical states, despite the results [3–6] of Seiji Terashima and co-authors that show how the duality is invalid for non-semiclassical states. Low-energy states are not necessarily semiclassical. (Section II)

- Terashima (2024) [6] offers how bulk-wise effective Hilbert space factorization differs significantly from that of the dual boundary theory. This has significance for the small corrections theorem, which relies on bulk-wise Hilbert space factorization between the black hole (BH) exterior and the interior, especially given that the example setup of Terashima (2024) has some particle-qubit resemblance with that of the small corrections theorem. (Section II)

- Assuming the entanglement wedge duality is valid in AdS/BCFT [10, 11] and states remain semiclassical, we apply the duality to AdS/BCFT and find a paradox, unless we deny bulk-wise Hilbert space factorization between brane $Q$ and standard boundary $M$. (Section III)

- A toy qudit (qubit) model is insufficient to disprove semiclassicality of black hole physics, since a slight change in Hilbert space factorization (even without complete breakdown in factorization) can restore unitarity without violating semiclassicality. We justify the slight change in light of the BCFT representation in double holography. (Section III D)

## II. ON TERASHIMA (2024)

### A. Validity of the entanglement wedge duality, restricted to semiclassical states

Terashima (2021), Terashima (2023) and Terashima (2024) [3, 5, 6] are studies of localized bulk states represented on the boundary CFT with regards to the entanglement wedge duality. The analysis shows that despite being localized in the bulk, such states generally require the whole space (full) support when represented as CFT states.

The main problem, unnoticed in these aforementioned papers, is that a bulk local state or an effectively local CFT state is not within the scope of bulk perturbation theory (semi-classical physics) that the results supporting the entanglement wedge duality, including the Jafferis-Lewkowycz-Maldacena-Suh (JLMS) relative entropy result [12], rely upon. We can understand some conditions required by bulk perturbation theory in terms of holographic entanglement entropy (HEE). HEE of region $A$ in CFT is typically stated as follows (with $G = c = \hbar = 1$) [12, 13]:

$$S_A = \frac{Area(\Sigma_A)}{4} + \text{s.c.} + \text{n.s.c.} \tag{1}$$

where $\Sigma_A$ is the extremal surface for $A$ in the bulk (with entanglement wedge being the sub-region enclosed by $A$ and $\Sigma_A$), s.c. refers to semiclassical corrections and n.s.c refers to non-semiclassical corrections. Bulk perturbation analysis then requires that:

$$\frac{Area(\Sigma_A)}{4} \gg |\text{s.c.}| \gg |\text{n.s.c}| \tag{2}$$

Bulk local states should have entanglement entropy close to zero for the supported bulk sub-region, which implies that Equation (2) breaks down: bulk perturbation theory becomes invalid. Similarly, an effectively localized CFT state also has entanglement entropy close to zero for the supported CFT region, which means that we cannot directly apply semiclassical tools to obtain the holographic bulk sub-region understanding.

States such that bulk perturbation theory remains valid, satisfying Equation (2), are referred to as semiclassical states. Terashima (2023) [5] says (quoted):

> "Note that in the free bulk theory limit, any state realized by smearing of a local operator as to set an energy cutoff can be regarded as a low-energy state.."

While this is technically true, such a low-energy bulk local state is not a semiclassical state: entanglement entropy deviates so much from the reference classical state (such as global vacuum) such that bulk perturbation analysis must break down. Conceptually, semiclassical states are not equivalent to low-energy states.

An argument can go that this still proves that at least some part of the entanglement wedge duality is incorrect, since applicability of the duality is then based on states, not operators. However, a more accurate interpretation would be that the reconstructed bulk field $\phi_{RC}(x)$ from the sub-region boundary-to-bulk map satisfies $\phi_{RC}(x) \approx \phi(x)$ within semiclassical physics. Similarly, while reconstructed $O_{\Delta}^{RC}$ (via the bulk-to-boundary map of the entanglement wedge reconstruction) is not a valid CFT operator, it can be used as if it is a CFT operator within semiclassical physics. This is made clearer by Equation 3.9 of Sugishita and Terashima (2022) [4], restated in this paper as Equation (3).

Unless we are interested in bulk physics that go beyond semiclassical bulk perturbation theory limits such as localized or unentangled states, the exact reconstruction of bulk fields is not necessary. This view is already affirmed in Almheiri et al. (2015) [1], where different reconstructed bulk fields (via the subregion boundary-to-bulk map) that should match to the same bulk field only differ in their non-semiclassical physics contents. (Again, see Equation (3).) A confusion may be there, since the paper uses word 'low-energy' - clearly, such low-energy states do not include low-energy localized states.

Sugishita and Terashima (2022) [4] - abbreviated as ST (2022) - adopts a slightly different strategy to disproving the entanglement wedge duality. Instead of bulk local states (and fields) on global AdS spacetime, we move onto bulk fields on the AdS-Rindler patch (AdS-

Rindler quantization). Equation 3.9 (the global CFT two-point function) in ST (2022) suggests how the naive bulk-to-boundary map remains effectively valid as a low-energy semiclassical approximation, replicated below:

$$\langle 0_{CFT}|T(O_\Delta^{CFT}(X)O_\Delta^{CFT}(X'))|0_{CFT}\rangle \approx e^{-\Delta(\Phi(X)+\Phi(X'))}\mathrm{Tr}_a\left[\rho_a T(O_\Delta^{RC}(X)O_\Delta^{RC}(X'))\right] \quad (3)$$

where $|0_{CFT}\rangle$ is global CFT vacuum, $O_\Delta^{CFT}$ is a global CFT operator, $a$ is a sub-region on the AdS-Rindler patch, $O_\Delta^{RC}$ is an AdS-Rindler-reconstructed bulk-to-boundary operator and $\rho_a = Tr_{\bar{a}}\left[|0_{bulk}\rangle\langle 0_{bulk}|\right]$, where $|0_{bulk}\rangle$ is global AdS vacuum. Equation (3) ('Equation 3.9') is similar to Equation 3.8 of ST (2022) for the same two-point function where actual sub-region CFT operators $O_\Delta^{CFT,flat}$ are used, and in this sense, $O_\Delta^{RC} \approx O_\Delta^{CFT,flat}$. This implies that pseudo-CFT operators constructed via the bulk-to-boundary map applied to the AdS-Rindler patch remain valid for semiclassical states, in contrast to the stated claim that the entanglement wedge duality is invalid - though in the end, this is an interpretational issue.

Equation 3.10 of ST (2022) says that for AdS-Rindler two-point functions:

$$\begin{gathered}\langle 0_a|T(O_\Delta^{RC}(X)O_\Delta^{RC}(X'))|0_a\rangle \not\approx \langle 0_A|T(O_\Delta^{CFT,flat}(X)O_\Delta^{CFT,flat}(X'))|0_A\rangle \\ \Rightarrow O_\Delta^{CFT,flat} \not\approx O_\Delta^{RC}\end{gathered} \quad (4)$$

where $|0_A\rangle$ is vacuum on boundary region $A$. However, note that sub-region vacuum $|0_a\rangle$ is not a semiclassical state from the global AdS point of view. Therefore, $O_\Delta^{CFT,flat} \not\approx O_\Delta^{RC}$ should be expected from the sub-region vacuum point of view.

### B. Bulk-wise effective Hilbert space factorization breakdown

Terashima (2024) [6] studies the wavepacket of Equation (5) and points out that the bulk theory and the dual CFT understand the wavepacket to have significantly different localization properties, which can be seen by computing CFT energy density $\mathcal{E}(x,t)$ of the wavepacket. Indeed, it is as if the CFT view has a particle branching off into two localized particles, when the bulk view has none of this, which seemingly makes the naive entanglement wedge duality impossible. The disproof strategy again comes from bulk localization, and we can again note that these localized states are not semiclassical states.

We state the key results of Terashima (2024). For the Poincaré patch of $\mathrm{AdS}_{d+1}$, the bulk wavepacket $|p, \bar{\omega}\rangle$ being studied goes as (assuming $a \ll 1$, $a^2 p^2 \gg 1$, $a\bar{\omega} \gg 1$):

$$|p, \bar{\omega}\rangle = \lim_{z \to 0} \frac{1}{z^{\Delta}} \int dt\, dx^i\, e^{-\frac{x^i x_i + t^2}{2a^2} + ip_i x^i - i\bar{\omega} t} \phi(t, z, x^i)|0\rangle \tag{5}$$

The localization property of the wavepacket in the bulk approximately goes as follows for the fixed $p, \bar{\omega}$ up to sufficient Gaussian suppression by small $a$:

$$\langle 0|\phi(t = 0, z, x^i) e^{iHt} |p, \bar{\omega}\rangle \propto e^{-\frac{1}{2a^2}\left((t \pm \bar{\omega} z/p_z)^2 + (x^j \mp p^j z/p_z)^2\right)} \tag{6}$$

where $p_z \equiv \sqrt{\bar{\omega}^2 - p^2}$. Since $z > 0$, for $t > 0$ this suggests localization around $z = \frac{p_z t}{\bar{\omega}}$, $x^j = -\frac{p^j t}{\bar{\omega}}$, confirming effectively one-particle localization.

Now assume $d = 2$. The localization property of the wavepacket in the CFT is then analyzed in terms of energy density function:

$$\mathcal{E}(t, x) \propto e^{-\frac{(x+t)^2}{2a^2}}(\bar{\omega} - p) + e^{-\frac{(x-t)^2}{2a^2}}(\bar{\omega} + p) \tag{7}$$

which suggests two-particle localization, with one particle localized around $x = -t$ and the other around $x = t$.

Worded differently, Terashima (2024) demonstrates that entirely different effective Hilbert space factorization could be there in the bulk relative to the dual CFT: effective one-particle (bulk) versus two-particle (CFT) Hilbert space factorization. This is interesting because in the black hole information problem, there is always a question about effective Hilbert space factorization. One popular idea is that the effective Hilbert space is not factorized into the black hole exterior and the interior, and information on the interior is actually on the exterior [14, 15]. Terashima (2024) suggests a possibility that while (especially vacuum-subtracted) Hilbert space factorization breaks down on the bulk, it does not on the dual boundary theory side: entanglement entropy of the black hole interior might be zero on the bulk side, while non-zero on the boundary theory side following the Page curve [16].

Note that the boundary CFT picture of the example in Terashima (2024) has single-qubit pair resemblance to the setup of the small corrections theorem, with two qubits (particles) flying away from each other and the horizon - see the left cylinder of Figure 1 for the boundary CFT representation of the Terashima (2024) example and the right cylinder of Figure 5 for the small corrections theorem setup. So without even explicitly introducing a black hole, we see some signs that echo Hilbert space factorization breakdown in the bulk: the quantum field theory on the boundary understands radiation pairs (a pair on the black

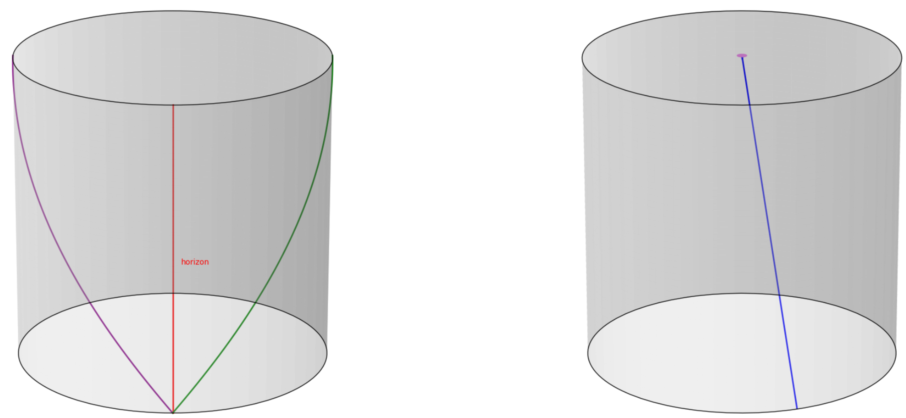


FIG. 1: AdS cylinder, as roughly relating to Terashima (2024). **Left**: two particle-like objects (purple and green curved lines) flying away from each other, as in the boundary representation of the example in Terashima (2024). The red line could be considered a horizon, if somewhat alluding to a black hole. **Right**: one particle-like object (blue line), as in the bulk representation of the example in Terashima (2024). The figure is rotated for visual clarity, but both the blue and the purple lines start from the same cylinder location. The blue line further reflects $x = -pt/\bar{\omega}$, moving away from the 'horizon'.

hole interior and the exterior) to have formed, while the quantum gravity theory on the bulk only has information on the black hole exterior. It would then make sense to ask for entanglement entropy of each 'particle' in the boundary theory, but not much so in the bulk, where two boundary theory particles are basically the same particle expressed differently.

This analogy does have limitations, since a black hole horizon is not located on AdS boundary. However, this has direct relevance when we discuss Karch-Randall braneworld, with brane $Q$ of the 'extended boundary' roughly working as (part of) the black hole interior.

We also note that the Hilbert space factorization issue suggests that a measurement theory in AdS/CFT can be complicated, though this complication may be a hint toward fully resolving the measurement problem (a final pure state measurement outcome against a pre-observation mixed state) For now, we leave this measurement theory question as mere speculation.

## III. ENTANGLEMENT WEDGE DUALITY (QUASI-PARADOX) ON KARCH-RANDALL BRANEWORLD

Having clarified conditions required for the entanglement wedge duality, we now discuss how the duality unfolds on Karch-Randall braneworld [17] and to lesser degree, AdS/BCFT [11]. In Section II, both the entanglement wedge duality and effective bulk-wise Hilbert space factorization broke down due to semiclassicality violations. By contrast, in Karch-Randall braneworld and especially in contexts of black hole physics, states in most circumstances are assumed to be semiclassical, and we expect the entanglement wedge duality to hold. The point of this section is to suggest via a quasi-paradox that this can only be possible if bulk-wise Hilbert space factorization between brane $Q$ and rest of bulk $N$ breaks down.

The state of affairs in Karch-Randall braneworld turns out to be fairly complicated with the four potentially overlapping possibilities, requiring us to sacrifice at least one of 1) bulk-wise Hilbert space factorization, 2) semiclassicality and 3) the entanglement wedge duality - the conventional choice would be to sacrifice the first option, which coincidentally also addresses the massive gravity criticism [7, 18] of Karch-Randall braneworld and the small corrections theorem [2].

- Scenario 1: In Karch-Randall braneworld, brane $Q$ features massive gravity regardless of a representation (bulk, brane, BCFT representations). The entanglement wedge duality and holographic entanglement entropy via the islands formula are preserved. See [7, 18] for this line of critiques, and see [19] for an analysis of AdS/BCFT with different boundary conditions (NBC, CBC, DBC) which also confirms massive gravity of brane $Q$ for all boundary conditions. [20] shows why massive gravity constrains applicability of Karch-Randall braneworld. In an alternative possibility, we evade these results by arguing that there is no Hilbert space factorization between brane $Q$ and rest of bulk $N$ in the bulk representation, just as the black hole information problem is sometimes avoided by denying Hilbert space factorization between the black hole interior and the exterior [15] - see Section III C as well.

- Scenario 2: Physics cannot be semiclassical in Karch-Randall braneworld, affirming the main message of the small corrections theorem [2] and, in some sense, Terashima (2024) in that the entanglement wedge duality breaks down once outside of semiclassical physics. It is natural to expect that holographic entanglement entropy (HEE) breaks down as well, though a slim chance exists such that HEE is saved despite breakdown of the duality.

- Scenario 3: The entanglement wedge duality is simply inapplicable in Karch-Randall braneworld, and arguments from AdS/CFT cannot be extrapolated to Karch-Randall braneworld, with the notion of semiclassicality in Karch-Randall braneworld different from that in conventional AdS/CFT.

- Scenario 4: Bulk-wise Hilbert space factorization between brane $Q$ and rest of bulk $N$ breaks down in spirit of Terashima (2024), saving the entanglement wedge duality. However, boundary-wise factorization between brane $Q$ and standard boundary $M$ is maintained. This addresses the massive gravity criticism of Karch-Randall braneworld and saves semiclassicality, while also conceptually denying applicability of the small corrections theorem [2]: in the bulk representation, there is no information on the black hole interior and brane $Q$ to begin with. In this understanding, the black hole interior in the bulk is something to be reconstructed, such as in the Papadodimas-Raju proposal [21].

The circumstance is further complicated by the fact that we do not know the exact quantum gravity theory on the bulk. We may expect the dual CFT (or BCFT) to provide a full probe into quantum gravity via AdS/CFT, but how operators and observables are related is yet to be reasonably understood.

After deriving the quasi-paradox in Section III A, we assume that states remain semi-classical and the entanglement wedge duality is safe to use, with holographic entanglement entropy calculations done via the islands formula, throughout Section III. This is justified in Section III C, which distinguishes validity of the entanglement wedge duality from that of bulk effective field theory. Furthermore, we assume NBC (Neumann boundary condition) as the boundary condition for brane $Q$ in Karch-Randall braneworld, but this choice is largely unimportant for the key arguments of this paper.

## A. A quasi-paradox

The quasi-paradox for the AdS/BCFT (and in general, Karch-Randall braneworld [17]) entanglement wedge duality is simple to demonstrate. In AdS/BCFT [11], there are instances where (bulk) extremal surface $\Sigma_R$ corresponding to CFT codimension-one region $R$ intersects with brane $Q$ and even surrounds parts of brane $Q$. See Figure 2 (the images replicated from [22]) that corresponds to a simple (Karch-Randall) black hole model in [22]. Entanglement wedge $N_R$ of region $R$ contains parts of brane $Q$, and in case the entanglement wedge duality holds, then semiclassical physics on $N_R \cap Q$ is captured by CFT semiclassical physics on region $R$, but this is paradoxical, unless bulk-wise Hilbert space factorization between bulk $N$ and brane $Q$ breaks down. This (quasi-)paradox assumes that boundary-wise Hilbert space factorization between standard AdS boundary $M$ and brane $Q$ remains valid, justified by BCFT considerations.

We note that this is a fairly general phenomenon in AdS/BCFT (and to full extent, Karch-Randall braneworld) as can be seen in Figure 3, which is consistent with the simple bulk action of Equation (8) [11].

One could further say that the AdS/BCFT quasi-paradox provides the non-overlapping boundary region variant of the quasi-paradox in [1] that motivated the quantum error correction view of bulk reconstruction. (The quasi-paradox of [1] is that a bulk operator far away from the boundary may be reconstructed from different overlapping boundary regions, but these reconstructed operators seem paradoxical, unless there are Hilbert space factorization differences between the bulk and the dual boundary theory.)

A naive critique to the above view may argue that entanglement wedge $N_R$ should exclude brane $Q$ - then both the entanglement wedge duality and bulk-wise Hilbert space factorization are maintained. But it is unclear what this even means, given that there are (re)action and backreaction between the brane and rest of the bulk. The islands formula, a modification of the standard holographic entanglement entropy calculation, further questions such a view, given that we incorporate the islands area term into entropy calculations - see Section III B for more discussions.

In the simplest example, the bulk action is given as [11]:

$$I_{bulk} = \frac{1}{16\pi G_{bulk}} \int_N \sqrt{-g}(R - 2\Lambda) + \frac{1}{8\pi G_{brane}} \int_Q \sqrt{-h}(K - T) \tag{8}$$

and so the action may be understood to consist of two sub-actions and sub-theories. However, the bulk theory is still one theory justified by the Randall-Sundrum mechanism [23] - it is just that at the classical effective level, it is as if there are two theories interacting.

### B. Islands and Hilbert space factorization breakdown

For now, we restrict the discussion of islands solely to Karch-Randall braneworld cases, where we rely on branes for island contributions instead of bulk corrections. Nevertheless, the discussion generalizes to the case where bulk corrections are utilized instead [24, 25].

There are multiple black hole-like implementations of Karch-Randall braneworld - this paper uses the implementation of [22] that sews two CFT copies with the Karch-Randall brane. Alternatively, there exists the black string model of [17], where two Karch-Randall braneworld copies (more precisely, two Karch-Randall brane copies) are connected by black string. Island contributions to holographic entanglement entropy are determined differently: in [22], we consider brane area contributions, whereas in [17], we compare candidate extremal surfaces that involve different boundary conditions and choose the one with minimal area - (standard) Hartman-Maldacena surfaces that do not intersect with Karch-Randall branes and pass through Einstein-Rosen bridges smoothly and 'island surfaces' that do not cross Einstein-Rosen bridges and end at the branes. Despite differing treatments of Karch-Randall branes, the lessons obtained in this subsection remain the same regardless of the models used, since the analysis is not tied to a particular model.

A modification of holographic entanglement entropy (HEE) with islands in the Karch-Randall braneworld model of [22] is given as:

$$S_R = \min\left\{\underset{\Sigma_R}{\mathrm{ext}}\left(\frac{Area(\Sigma_R)}{4G_{bulk}} + \frac{Area(\sigma_R)}{4G_{brane}}\right)\right\} \tag{9}$$

with islands $\sigma_R = \Sigma_R \cap Q$. Intuitively, the formula considers AdS/CFT applied to the BCFT boundary (and the brane) and incorporates the leading area term contribution from brane-side AdS/CFT to $S_R$. This intuition contradicts the idea that the entanglement wedge should leave out brane $Q$ parts. As an example, one can refer to Figure 2.

Even when the standard RT HEE calculation (without islands) is invoked, as with the setup of [11] (with the simple bulk action described in Equation (8)) it is possible that $\Sigma_A$ (for CFT region $A$) intersects with brane $Q$ - see Figure 3, which means that the quasi-paradox applies fairly generally in AdS/BCFT [11] and Karch-Randall braneworld [17, 23].

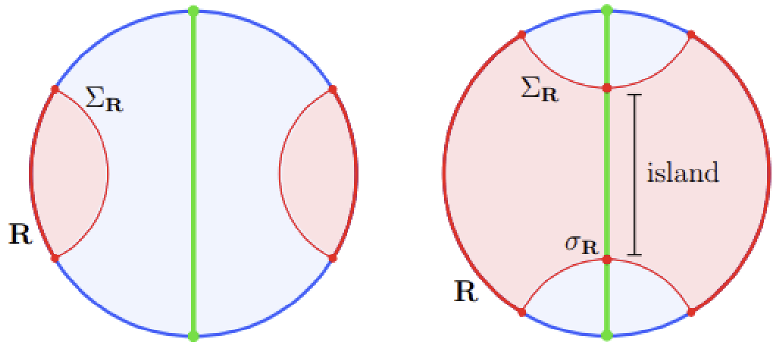


FIG. 2: In the bulk Karch-Randall braneworld representation of [22]. The circle represents standard boundary $M$ in the bulk, and the interior of the circle is bulk $N$. For $M$, we now have two identical CFT copies (thermofield double state formalism [26]) pieced together by brane $Q$ in the center line. $R$ is on both the left CFT and the right CFT. For the right figure, extremal surface $\Sigma_R$ touches $Q$, which creates islands $\sigma_R$ on the brane. The point is that in such a case, we need to incorporate $\sigma_R$ to holographic entanglement entropy (HEE) calculations of $R$. The image is from [22], distributed with license CC-BY 4.0 (http://creativecommons.org/licenses/by/4.0/) which permits any use, distribution and reproduction in any medium, provided the original authors and source are credited.

Therefore, unless we give up 1) semiclassicality or 2) the entanglement wedge duality, there is bulk-wise Hilbert space factorization breakdown in Karch-Randall braneworld such that:

$$\begin{gathered}\mathcal{H} = \mathcal{H}_M \otimes \mathcal{H}_Q \\ \mathcal{H} \neq \mathcal{H}_{N-Q} \otimes \mathcal{H}_Q \\ \mathcal{H} \neq \mathcal{H}_{BH} \otimes \mathcal{H}_{N-BH}\end{gathered} \tag{10}$$

where $\mathcal{H}_M$ refers to the Hilbert space of standard boundary $M$, $\mathcal{H}_{BH}$ refers to the Hilbert space of the black hole interior and so forth, with $N$ being bulk spacetime. This would imply that the small corrections theorem [2] is inapplicable, since it relies on bulk-wise Hilbert space factorization between the BH interior and the exterior.

We briefly note the less straightforward example: in the black string model of [17], brane contributions to entanglement entropy arise from the fact that the entanglement wedge of two connected standard boundary surface copies bound by the island surface involves two bulk copies connected by black string - in order for two copies to be connected, information

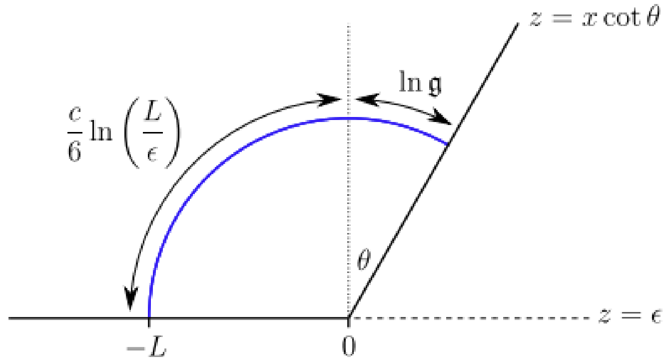


FIG. 3: In the AdS/BCFT bulk representation of [11]. We can consider the line from $x = 0$ to $x = -L$ to be region $A$ of standard boundary $M$. $\Sigma_A$ is the curved blue line in the figure, hitting $Q$ that intersects with $M$ at angle of $\theta$. Holographic entanglement entropy ceases to be dominated by the non-brane area (length) term as $\ell \to 0$ with $\ell = L$, which is the point made in [11]. The image is from [27], distributed with license CC-BY 4.0 (http://creativecommons.org/licenses/by/4.0/) which permits any use, distribution and reproduction in any medium, provided the original authors and source are credited.

on the brane cannot be discarded. See Figure 4.

### C. Validity of effective field theory is a different question from validity of the entanglement wedge duality

Breakdown of bulk-wise Hilbert space factorization suggests that bulk effective field theory (EFT) is modified, even when background bulk metric remains the same. This is indeed expected in black hole physics, since the very question of the black hole information problem is that bulk EFT states are clearly invalid at late-time evaporation.

Bulk EFT correction is an expected result from Sugishita and Terashima (2022) [4] as well, whenever AdS space is modified or deformed (such as introducing cutoff surfaces or taking the Rindler-AdS patch as in [4]). Indeed, the lesson there was that low-energy bulk states in such modified AdS space do not necessarily have dual low-energy boundary theory states (even though inverse is necessarily so), and tachyonic CFT modes are required to represent some low-energy bulk states.

This is consistent with the result of [28], which attempts to resolve the causality para-

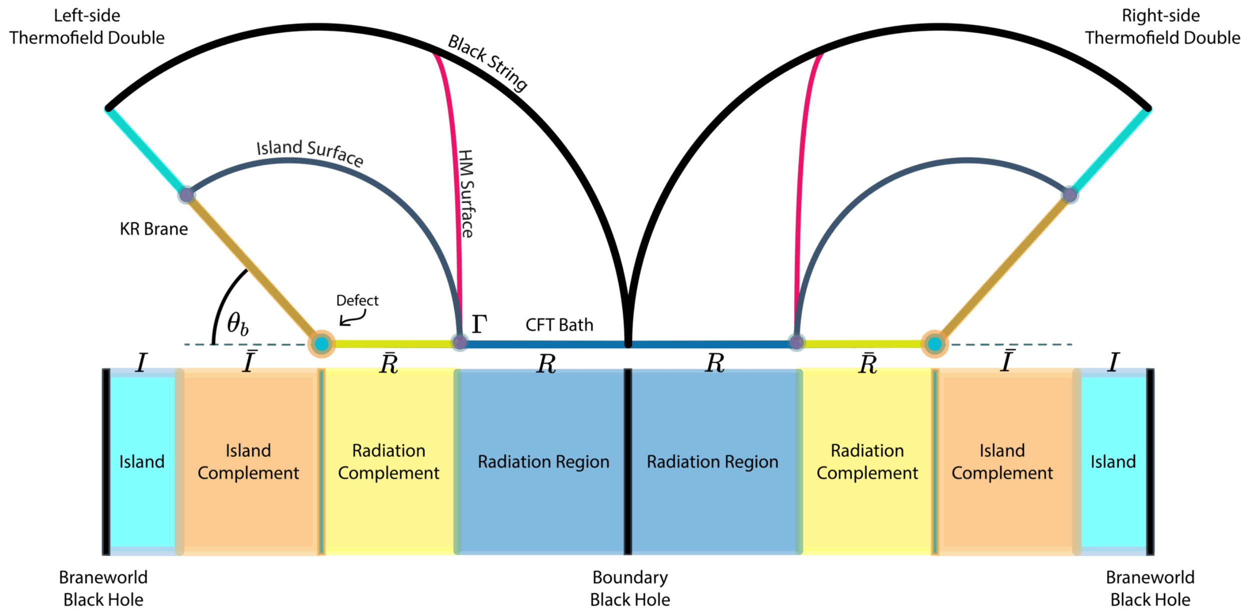


FIG. 4: The pictorial representation of the planar black string model and late-time islands in [17]. The Hartman-Maldacena (HM) surface (standard extremal surface) starts to be larger than the island surface, and the entanglement entropy of radiations starts to be governed by the island surface. The image is from [17], distributed with license CC-BY 4.0 (http://creativecommons.org/licenses/by/4.0/) which permits any use, distribution and reproduction in any medium, provided the original authors and source are credited.

dox in double holography of Karch-Randall braneworld. In particular, [28] confirms that EFT domain of dependence is different from holographic domain of dependence - in short, validity of bulk EFT for the entanglement wedge is a different question from validity of the entanglement wedge duality itself. Even without reliable bulk EFT, the boundary theory can be relied upon to make correct low-energy predictions, aided by the entanglement wedge duality.

The aforementioned difference between the entanglement wedge duality and EFT is important in clarifying the massive versus massless gravity issue [29] in Karch-Randall braneworld. As an approximation, there are reasonable instances where coupled brane EFT remains valid, which is what we expect for analogous early-time evaporation. However, it is only reliable as an initial approximation, and the 'bulk theory' for the 'brane' is eventually significantly corrected. The question of massive gravity versus massless gravity on the brane is therefore difficult to address purely at the EFT level, unless truly quantum gravity analysis are invoked. Some holographic bulk EFT can be used to define the dual CFT via the EFT ('weak') version of holography, but holographic bulk EFT in Karch-Randall braneworld eventually has to be corrected if we trust the low-energy understanding of the CFT.

### D. A toy Karch-Randall braneworld qudit model of black hole evaporation

We now simplify Karch-Randall braneworld to a toy qudit model to describe black hole evaporation, with some losses as would be inevitable for a toy qudit model, grounded upon bulk-wise Hilbert space factorization breakdown between brane $Q$ and standard boundary $M$. While a black hole in Karch-Randall braneworld is usually described in the thermofield double formalism [26], we utilize a one-sided black hole in this toy qudit model.

The toy qudit model here suggests that whenever we have a representation similar to the BCFT representation of Karch-Randall braneworld (see the left cylinder of Figure 5), the conclusion of the small corrections theorem [2] is reversed despite having essentially the same semiclassical physics due to there only being one interior qudit. In this circumstance, we could think of the horizon (or BCFT boundary) as carrying information about the black hole (BH) interior in fashion of black hole complementarity [30].

#### 1. *Toy qudit model setup*

*a. Visual representation* The setup of the toy qudit model is pictorially described in the left cylinder of Figure 5. There are $N$ exterior qudits and one interior qudit (red line, horizon), which contrasts with the setup of the small corrections theorem where there are total of $N$ exterior qubits and $N$ interior qubits.

*b. Radiation pair Hamiltonian* The old (stale and only) interior qudit keeps interacting with new exterior qudits, each with interaction duration $\Delta t$, with Hamiltonian $H$ of each radiation pair around the horizon for interaction duration $\Delta t$ given by

$$H = H_{ext} + H_{BH} + H_I \tag{11}$$

with $H_{ext}$ being the non-interaction Hamiltonian of a new exterior qudit (purple lines in Figure 5), $H_{BH}$ being the non-interaction Hamiltonian of the interior qudit (the red line for the left cylinder of Figure 5) and $H_I$ is the interaction Hamiltonian between the interior qudit and the new exterior qudit.

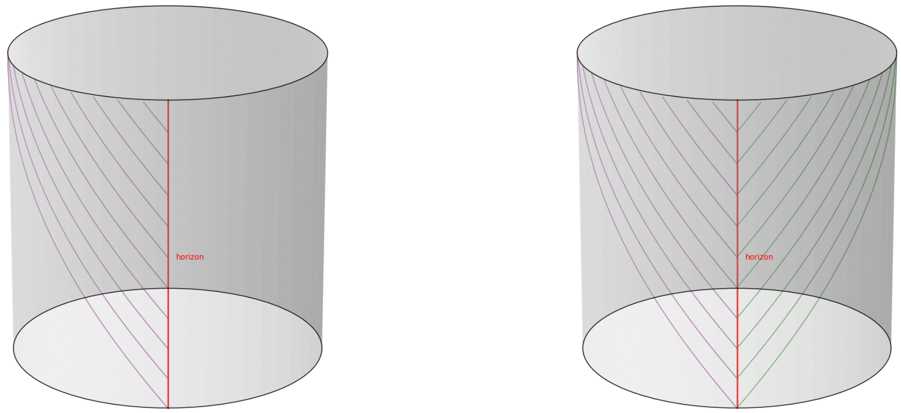


FIG. 5: The qudits in question live on the cylinder surface with time flowing upward. **Left**: the qudit setup of the toy qudit model in this paper. The left purple lines represent the black hole (BH) exterior qudits, moving away from the horizon. These qudits have effective Hamiltonian $H_{ext}$, at least while they interact with the BH interior qudit. The red line represents the sole BH interior qudit (that can also be seen as representing the horizon), with effective Hamiltonian $H_{BH}$. Each qudit pair (one on the exterior and the other on the interior) interacts for time $\Delta t$, with interaction Hamiltonian $H_I$. The fresh exterior qudits on the horizon are at the ground state of $H_{ext}$, waiting to extract energy from the BH and then move away from the horizon. **Right**: The toy qubit model setup of the small corrections theorem [2], with the BH interior qudit represented as green lines, separate from the horizon marked as the red line. Both images are only rough visual representations.

*c. The initial state of the new exterior qudit and the interaction setup* For the new exterior qudit at $t = k\Delta t$ (with $k \in \mathbb{N}$), its initial state is $|0_{ext}\rangle$, defined as the unique ground state of $H_{ext}$. The new exterior qudit at time $t = k\Delta t$ then interacts with the interior qudit for time interval of $\Delta t$ and then escapes away from the horizon, never interacting again with the interior qudit.

*d. The time-zero state of the first radiation pair* At time $t = 0$, the initial state of the first radiation pair is:

$$|\Psi(0)\rangle = |0_{ext}\rangle|\psi_{BH,0}\rangle \tag{12}$$

with $|\psi_{BH,0}\rangle$ being some **non-ground** state of $H_{BH}$.

*e. A constraint on interaction Hamiltonian $H_I$* Let $U = e^{-iH\Delta t}$ (with $\hbar = 1$). Then we constrain $H_I$ to satisfy the following entanglement structure with $\langle\psi_{BH,a}|\psi_{BH,b}\rangle = 0$, $\langle a_{ext}|b_{ext}\rangle = 0$:

$$U\left(|0_{ext}\rangle|\psi_{BH,0}\rangle\right) = \frac{1}{\sqrt{2}}\left(|a_{ext}\rangle|\psi_{BH,a}\rangle + |b_{ext}\rangle|\psi_{BH,b}\rangle\right) \tag{13}$$

This means that we are on the same page with the small corrections theorem of [2], entanglement-wise. The same semiclassical physics is shared with the small corrections theorem, except that there is a lack of the interior qudits in our setup. This allows us to reverse the conclusion of the small corrections theorem and demonstrate final purity of the BH exterior.

For describing early time interaction effects for the exterior qudits, we assume that the mixed state of the BH interior may be approximated by $|\psi_{BH,0}\rangle$, though the approximation breaks down as we move away form early times.

*f. Quantum channel $\phi$ from $U$* Restricted to the interior qudit, we can understand effects of $U = e^{-iH\Delta t}$ as applying quantum channel $\phi$ of time duration $\Delta t$, defined as:

$$\phi(\rho_{BH}) \equiv \phi(|0_{ext}\rangle\langle 0_{ext}|, \rho_{BH}) \equiv \mathrm{Tr}_{ext}\left[U\left(|0_{ext}\rangle\langle 0_{ext}| \otimes \rho_{BH}\right)U^{\dagger}\right] \tag{14}$$

with $\rho_{BH}$ being the mixed state of $BH$, and $\mathrm{Tr}_{ext}$ is partial trace over the BH exterior qudit the interior qudit is interacting with. Assume that $\phi$ is a quantum thermal operation as defined in [31], which is about non-interaction energy conservation:

$$[U, H_{ext} + H_{BH}] = 0 \tag{15}$$

If $\rho_{BH} = |0_{BH}\rangle\langle 0_{BH}|$, with $|0_{BH}\rangle$ being the unique ground state of $H_{BH}$ ($|0_{BH}\rangle \neq |\psi_{BH,0}\rangle$), then Equation (15) with the BH exterior qudit being $|0_{ext}\rangle$ implies that $|0_{BH}\rangle\langle 0_{BH}|$ is a fixed point of channel $\phi$ due to both exterior and interior qudits stuck at ground energies that cannot be lowered further. Let this fixed point be the unique one for channel $\phi$.

### 2. Demonstration of final purity of the exterior qudits

We now show that Equation (16) is satisfied:

$$\lim_{n\to\infty}\phi^n(\rho_{BH}) \equiv \lim_{n\to\infty}\underbrace{\phi\circ\phi\cdots\phi}_{n}(\rho_{BH}) = |0_{BH}\rangle\langle 0_{BH}| \tag{16}$$

*Proof.* Since the exterior qudit is initially at ground state $|0_{ext}\rangle$ and none of excited states of the interior qudit are fixed points of $\phi$, the interior qudit in initially excited non-ground states results in non-static evolution and the interior qudit has to lose energy. Therefore, as $\phi$ is applied infinite times, the interior qudit has to lose all energy and collapse into ground state $|0_{BH}\rangle$, which is the fixed point of channel $\phi$. $\square$

As a result, as $t \to \infty$, exterior qudits together are purified (due to the interior qudit being purified), with unitarity restored.

### 3. *Final holographic steps*

Having established purity in the toy BCFT qudit model, AdS/BCFT intuitions are invoked to justify the bulk representation of black hole evaporation in case the toy qudit model dutifully captures relevant BCFT physics. The understanding is that bulk-wise Hilbert space factorization breaks down and there is actually nothing inside the black hole interior information-wise in the bulk representation, differing from the BCFT representation.

Finally, we note that the toy qudit model can more specifically be viewed as simplification of the moving mirror AdS/BCFT models, with review provided in [8, 9].

### 4. *A general lesson, outside of Karch-Randall braneworld*

Fundamentally, the reason why we can avoid the small corrections theorem in the toy qudit model of holography is fairly general, not dependent on particular holographic setups. We only require the BH interior qudit to remain on the horizon as in Figure 5. The horizon can then be taken as the boundary of BCFT, but this is an additional choice made to utilize AdS/BCFT.

It is this horizon structure that gives rise to purification of the BH exterior over time in the toy qudit model, largely consistent with black hole complementarity [30] - in one sub-representation that emphasizes the black hole interior, information passes through the horizon, while in an alternative sub-representation, information resides on the horizon, continuing to interact with the BH exterior.

## IV. CONCLUSION

We now review the paper in reverse. The toy holographic qudit model was developed in Section III D, which suggests that as long as there is a representation of the universe akin to the BCFT representation of AdS/BCFT that has black hole interior information residing on some horizon (in fashion of black hole complementarity), then purification in black hole evaporation can be demonstrated.

When bulk semiclassicality is satisfied, the entanglement wedge duality is assumed to be satisfied. However, the information quasi-paradox of Section III suggests we have to give up one of the three: 1) bulk-wise Hilbert space factorization between brane $Q$ and rest of the bulk $N$, 2) states being semiclassical, 3) entanglement wedge duality. We give up the first to save the entanglement wedge duality, which also provides why the small corrections theorem does not apply - in the bulk, Hilbert space factorization breaks down between the black hole interior and the exterior. We can choose to factorize the Hilbert space in a way that semiclassical physics around the black hole horizon is emphasized, but this leads to the interior and the exterior being mixed from the Hilbert space factorization point of view, which means entanglement entropy results are unreliable. There is no breakdown in semiclassical physics around the horizon (though corrections to bulk effective field theory are there), only that the Hilbert space is not what is initially implied by the pre-quantized classical theory.

It is reasonable to ask how breakdown in bulk-wise Hilbert space factorization can further be justified. Terashima (2024) [6] provides an excellent example in AdS/CFT on how this can occur, which is replicated in Section II. There are two localized particles moving away from each other (in the $x$ direction) in the CFT representation, but only one localized particle in the bulk representation, moving in the negative $x$ direction toward the radial center. Even without a black hole or an additional brane, such breakdown is general. The example of Terashima (2024) is not semiclassical, and the entanglement wedge duality does not apply a priori. Before a series of papers by Seiji Terashima [3–6], we could have hoped that the entanglement wedge duality applies beyond semiclassical cases. This turns out to be impossible, emphasizing how tied the duality is to semiclassical conditions.

For now, the chance of the quasi-paradox of Section III actually suggesting the impossibility of semiclassical Karch-Randall braneworld states remains. In such a case, the toy qudit model of Section III D cannot be used to argue against the small corrections theorem, since the quasi-paradox actually vindicates impossibility of semiclassicality. Bulk-wise Hilbert space factorization breakdown can be used to avoid the massive gravity criticisms [7, 18] against Karch-Randall braneworld (see Section III C as well), but without such breakdown, massive gravity haunts Karch-Randall braneworld and its logical consistency within even minimal empirical requirements can be questioned [20]. Future works should clarify these issues.

## DATA AVAILABILITY AND DECLARATION OF INTERESTS

No data were utilized in this theoretical paper. The author(s) have no funding source to declare. Furthermore, there is no conflict of interests.

---

[1] Ahmed Almheiri, Xi Dong, and Daniel Harlow, "Bulk locality and quantum error correction in AdS/CFT," Journal of High Energy Physics **2015** (2015), 10.1007/jhep04(2015)163.

[2] Samir D. Mathur, "The information paradox: a pedagogical introduction," Classical and Quantum Gravity **26**, 224001 (2009).

[3] Seiji Terashima, "Bulk locality in the AdS/CFT correspondence," Physical Review D **104** (2021), 10.1103/physrevd.104.086014.

[4] Sotaro Sugishita and Seiji Terashima, "Rindler bulk reconstruction and subregion duality in AdS/CFT," Journal of High Energy Physics **2022** (2022), 10.1007/jhep11(2022)041.

[5] Seiji Terashima, "Simple bulk reconstruction in anti-de Sitter/conformal field theory correspondence," Progress of Theoretical and Experimental Physics **2023** (2023), 10.1093/ptep/ptad054.

[6] Seiji Terashima, "Wave packets in AdS/CFT correspondence," Phys. Rev. D **109** (2024), 10.1103/PhysRevD.109.106012.

[7] Hao Geng, Andreas Karch, Carlos Perez-Pardavila, Suvrat Raju, Lisa Randall, Marcos Riojas, and Sanjit Shashi, "Information transfer with a gravitating bath," SciPost Physics **10** (2021), 10.21468/scipostphys.10.5.103.

[8] Ibrahim Akal, Yuya Kusuki, Noburo Shiba, Tadashi Takayanagi, and Zixia Wei, "Entanglement Entropy in a Holographic Moving Mirror and the Page Curve," Phys. Rev. Lett. **126** (2021), 10.1103/PhysRevLett.126.061604.

[9] Ibrahim Akal, Yuya Kusuki, Noburo Shiba, Tadashi Takayanagi, and Zixia Wei, "Holographic moving mirrors," Classical and Quantum Gravity **38** (2021), 10.1088/1361-6382/ac2c1b.

[10] Mitsutoshi Fujita, Tadashi Takayanagi, and Erik Tonni, "Aspects of ads/bcft," Journal of High Energy Physics **2011** (2011), 10.1007/jhep11(2011)043.

[11] Tadashi Takayanagi, "Holographic dual of a boundary conformal field theory," Physical Review Letters **107** (2011), 10.1103/physrevlett.107.101602.

[12] Daniel L. Jafferis, Aitor Lewkowycz, Juan Maldacena, and S. Josephine Suh, "Relative entropy equals bulk relative entropy," Journal of High Energy Physics **2016** (2016), 10.1007/jhep06(2016)004.

[13] Shinsei Ryu and Tadashi Takayanagi, "Holographic Derivation of Entanglement Entropy from the anti–de Sitter Space/Conformal Field Theory Correspondence," Physical Review Letters **96** (2006), 10.1103/physrevlett.96.181602.

[14] Suvrat Raju, "Lessons from the information paradox," Physics Reports **943**, 1–80 (2022).

[15] Suvrat Raju, "Failure of the split property in gravity and the information paradox," Classical and Quantum Gravity **39** (2022), 10.1088/1361-6382/ac482b.

[16] Don N. Page, "Information in black hole radiation," Physical Review Letters **71** (1993), 10.1103/physrevlett.71.3743.

[17] Hao Geng, Andreas Karch, Carlos Perez-Pardavila, Suvrat Raju, Lisa Randall, Marcos Riojas, and Sanjit Shashi, "Entanglement phase structure of a holographic BCFT in a black hole background," Journal of High Energy Physics **2022** (2022), 10.1007/jhep05(2022)153.

[18] Hao Geng and Andreas Karch, "Massive islands," Journal of High Energy Physics **2020** (2020), 10.1007/jhep09(2020)121.

[19] Chong-Sun Chu and Rong-Xin Miao, "Conformal boundary condition and massive gravitons in AdS/BCFT," Journal of High Energy Physics **2022** (2022), 10.1007/jhep01(2022)084.

[20] Brando Bellazzini, Giulia Isabella, Sergio Ricossa, and Francesco Riva, "Massive gravity is not positive," Physical Review D **109** (2024), 10.1103/physrevd.109.024051.

[21] Kyriakos Papadodimas and Suvrat Raju, "An infalling observer in AdS/CFT," Journal of High Energy Physics **2013** (2013), 10.1007/jhep10(2013)212.

[22] Hong Zhe Chen, Robert C. Myers, Dominik Neuenfeld, Ignacio A. Reyes, and Joshua Sandor, "Quantum extremal islands made easy. Part II. Black holes on the brane," Journal of High Energy Physics **2020** (2020), 10.1007/JHEP12(2020)025.

[23] Hong Zhe Chen, Robert C. Myers, Dominik Neuenfeld, Ignacio A. Reyes, and Joshua Sandor, "Quantum extremal islands made easy. Part I. Entanglement on the brane," Journal of High Energy Physics **2020** (2020), 10.1007/JHEP10(2020)166.

[24] Ahmed Almheiri, Raghu Mahajan, Juan Maldacena, and Ying Zhao, "The Page curve of Hawking radiation from semiclassical geometry," Journal of High Energy Physics **2020**, 149 (2020).

[25] Geoffrey Penington, "Entanglement wedge reconstruction and the information paradox," Journal of High Energy Physics **2020** (2020), 10.1007/JHEP09(2020)002.

[26] Juan Maldacena, "Eternal black holes in anti-de Sitter," Journal of High Energy Physics **2003** (2003), 10.1088/1126-6708/2003/04/021.

[27] Seamus Fallows and Simon F. Ross, "Islands and mixed states in closed universes," Journal of High Energy Physics **2021** (2021), 10.1007/jhep07(2021)022.

[28] Dominik Neuenfeld and Manu Srivastava, "On the causality paradox and the Karch-Randall braneworld as an EFT," Journal of High Energy Physics **2023** (2023), 10.1007/JHEP10(2023)164.

[29] Hao Geng, Andreas Karch, Carlos Perez-Pardavila, Lisa Randall, Marcos Riojas, Sanjit Shashi, and Merna Youssef, "Constraining braneworlds with entanglement entropy," SciPost Physics **15** (2023), 10.21468/scipostphys.15.5.199.

[30] Leonard Susskind, Lárus Thorlacius, and John Uglum, "The stretched horizon and black hole complementarity," Physical Review D **48**, 3743–3761 (1993).

[31] Philippe Faist, Mario Berta, and Fernando G. S. L. Brandao, "Thermodynamic implementations of quantum processes," Communications in Mathematical Physics **384**, 1709–1750 (2021).