

How to count states in gravity

Vijay Balasubramanian^{1,2,3,*} and Tom Yildirim^{4,†}

¹*Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA*

²*Theoretische Natuurkunde, Vrije Universiteit Brussel and International Solvay Institutes, Pleinlaan 2, B-1050 Brussels, Belgium*

³*Rudolf Peierls Centre for Theoretical Physics, University of Oxford, Oxford OX1 3PU, United Kingdom*

⁴*Department of Physics, Keble Road, University of Oxford, Oxford OX1 3RH, United Kingdom*



(Received 5 March 2026; accepted 26 June 2026; published 23 July 2026)

Gibbons and Hawking proposed that the Euclidean gravity path integral with periodic boundary conditions in time computes the thermal partition sum of gravity. As a corollary, they argued that a derivative of the associated free energy with respect to the Euclidean time period computes gravitational entropy. Why is this interpretation correct? That is, why does this path integral compute a trace over the Hilbert space? Here, we show that the quantity computed by the Gibbons-Hawking path integral is equal to an *a priori* different object—an explicit thermal trace over the Hilbert space spanned by states produced by the Euclidean gravity path integral. This follows in two ways; (a) if the Hilbert space with two boundaries factorizes into a product of two single boundary Hilbert spaces, as we have previously shown, and (b) via explicit resolution of the trace by a spanning basis of states. We similarly show how a replicated Euclidean gravity path integral with a single periodic boundary computes a Hilbert space trace of powers of the density matrix, explaining why this approach computes the entropy of states entangled between two universes.

DOI: [10.1103/9v3t-91qh](https://doi.org/10.1103/9v3t-91qh)

I. INTRODUCTION

The Bekenstein-Hawking entropy formula $S_{\text{BH}} = \frac{A}{4G_N}$ is believed to universally relate the number of microstates of a black hole to its horizon area A . This formula was proposed through an analogy between classical laws of black hole mechanics and the laws of thermodynamics [1,2]. The analogy was made literal by Hawking's result [3] that black holes evaporate by emitting (seemingly) thermal radiation at temperature $T_H = \frac{1}{8\pi M_{\text{GN}}}$, derived by considering quantum field theory on black hole backgrounds. Thus, it appears that black holes are finite-dimensional quantum systems with Hilbert spaces of dimension $e^{S_{\text{BH}}}$.

A second argument to the same effect was proposed by Gibbons and Hawking [4]. They suggested that the Euclidean gravitational path integral with periodic boundary conditions in time should be interpreted as the thermal partition sum Z of gravity. This is in analogy with statistical

field theory where the partition sum is, by construction, computed by the Euclidean path integral with a periodic time β equal to the inverse temperature. Leaving aside some technical subtleties, the leading saddle point of the Gibbons-Hawking path integral turns out to be the Euclidean black hole, and so in the semiclassical limit $Z \approx e^{-I_{\text{BH}}}$ where I_{BH} is the black hole action. Pursuing the analogy with thermodynamics, we infer that the entropy is $S = (1 - \beta \partial_\beta) \ln Z$, where β is the Euclidean time period. This computation reproduces the Bekenstein-Hawking entropy formula.

But does the Gibbons-Hawking path integral actually compute a thermal trace over the Hilbert space? In statistical physics the path integral computing this trace is performed over a cylinder, i.e., a manifold of the form $\mathcal{M} \times S^1$ where the period of the circle is the inverse temperature β . But in gravity we must sum over all topologies and geometries that are asymptotic to $\partial\mathcal{M} \times S^1$. In fact, in the Euclidean black hole saddle point which dominates the path integral in appropriate regimes, the S^1 is contractible in the interior—the topology is that of a disc, not a cylinder. To take the $\partial/\partial\beta$ derivative required for computing the entropy we have to vary the period, and if we do this while keeping the rest of the saddle-point metric fixed, a conical singularity develops at the origin. This off shell geometry then has the topology of a cylinder, but it is

*Contact author: vijay@physics.upenn.edu

†Contact author: tom.yildirim@physics.ox.ac.uk

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) 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.

not clear why this is sufficient to admit a trace interpretation of the path integral. More generally, it is not clear whether the Gibbons-Hawking path integral computes a trace over some Hilbert space, what this Hilbert space might be, and whether this Hilbert space spans the possible states of quantum gravity.

Here, we establish that the Gibbons-Hawking path integral equals the thermal trace $\text{Tr}_{\mathcal{H}_\chi}(e^{-\beta H})$ over the Hilbert space $\mathcal{H}_\chi$ of quantum gravity with a single boundary. We show this in two ways by using the methods of [5]. First, we show that the result follows if the Hilbert space of quantum gravity with two boundaries factorizes into a product of Hilbert spaces with one boundary, as we have recently demonstrated [6]. In detail, we strategically insert a novel resolution of the identity in the two-boundary gravity Hilbert space into the Euclidean path integral with periodic time [5]. The terms in the resulting sum are equal to the terms in the thermal trace over the single boundary Hilbert space. Second, we use a tractable basis of states for the single boundary Hilbert space [6] to explicitly evaluate the path integrals in the thermal trace, and show that their sum collapses to a Euclidean path integral with a single periodic boundary.¹ In the regime of parameters where the leading saddle point of the Gibbons-Hawking path integral is a Euclidean black hole, these calculations give a micro-canonical Hilbert space dimension of $e^{A/4GN}$ where A is the horizon area after Lorentzian continuation of the saddle point. However, as we will discuss, the basis of states explaining this dimension need not have horizons at all.

We then consider states of universes with two asymptotic boundaries. Since the Hilbert space of such universes is a product of factors [6], we can determine the reduced density matrix on either factor by tracing out the other. The Rényi entropies should then be computed as a trace of powers of this density matrix, and the von Neumann entropy is recovered as a limit. However, the conventional approach instead computes a replicated Euclidean gravity path integral with a single periodic boundary, by analogy with the Gibbons-Hawking proposal for the gravitational partition sum. We show why this replicated integral is equal to the sum over path integrals that calculates the Hilbert space trace required for the Rényi entropy.

Four sections follow. In Sec. II we review how bases of states can be constructed by cutting open the gravitational path integral. In Sec. III we show why the Gibbons-Hawking path integral computes a thermal trace over the Hilbert space. In Sec. IV we explain why a replica path integral can evaluate the trace required to compute gravitational entanglement entropy. Finally, we conclude with a discussion in Sec. V.

¹These single-sided shell states are a single-boundary version of the two-sided shell states used to provide a microscopic understanding of black hole entropy in [7–11]. They were described in [12,13].

II. STATES FROM THE GRAVITATIONAL PATH INTEGRAL

We would like to compute the thermal partition function of a gravity theory obtained by tracing over an orthonormal basis of states $\text{Tr}_{\mathcal{H}_\chi}(e^{-\beta H}) = \sum_\gamma \langle v_\gamma | e^{-\beta H} | v_\gamma \rangle$. We need three ingredients: (1) A spanning set of states which could be overcomplete; (2) The action of the Hamiltonian on these states; (3) A way of taking the trace over the Hilbert space. As we will explain, all three are naturally available for states defined by cutting open the Euclidean gravitational path integral.

How to make states. First we review the construction of states by cutting open the Euclidean gravitational path integral, understood as an effective description below a cutoff, completed at high energy by some fundamental theory. See [5,6] for details. Our considerations will apply to universes with either vanishing or negative cosmological constant. The path integral computes a sum over topologies and geometries (g), and matter fields (ϕ), compatible with specified asymptotic boundary conditions (boundary topology, metric, and matter sources, collectively denoted by Σ_b): $\zeta[\Sigma_b] = \int_{g,\phi \rightarrow \Sigma_b} \mathcal{D}g \mathcal{D}\phi e^{-I_{\text{grav}}[g,\phi]}$, where I_{grav} is the gravitational and matter action. To construct states we partition the boundary $\mathcal{M}_b = \mathcal{M}_1 \cup \mathcal{M}_2$ so that $\partial\mathcal{M}_1 = \partial\mathcal{M}_2 = \mathcal{X}$ (Fig. 1). The resulting functional of the fields and geometry on the cut defines states $|\mathcal{M}_{1,2}\rangle \in \mathcal{H}_\chi$ in a Hilbert space $\mathcal{H}_\chi$ such that $\langle \mathcal{M}_2 | \mathcal{M}_1 \rangle = \zeta[\Sigma_b]$. We generally resort to saddle-point approximations to evaluate overlaps of these states. Given a cut $\mathcal{X}$ of a boundary manifold $\mathcal{M}_b$ we can produce different states $|i\rangle$ by inserting matter operators $\mathcal{O}_i$ on $\mathcal{M}_b$. To compute overlaps $\langle i | j \rangle$ we evaluate the path integral with $\mathcal{O}_{i,j}$ inserted on $\mathcal{M}_{1,2}$. The bra $\langle i |$ associated to the ket $|i\rangle$ defined by inserting $\mathcal{O}_i$ on $\mathcal{M}_1$ is obtained by inserting $\mathcal{O}_i^\dagger$ on $\mathcal{M}_1$. The ket $|i^*\rangle$ defined inserting $\mathcal{O}_i^\dagger$ on $\mathcal{M}_1$ therefore defines the same boundary condition as $\langle i |$, and so $\langle i | j \rangle = \langle j^* | i^* \rangle$. Finally, the leading saddle point of the path integral for the norm of a state can be sliced open at a moment time-reflection symmetry, provided this exists, and analytically continued to Lorentzian signature. This procedure associates a Lorentzian geometry to the state.

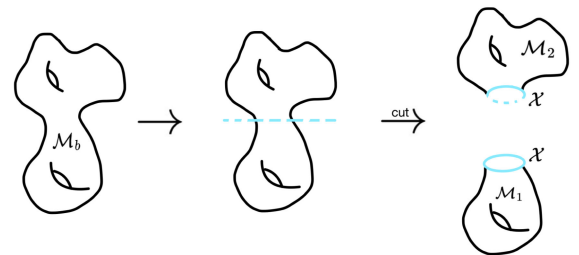


FIG. 1. Slicing a gravity path integral to define a state (adapted from [5]).

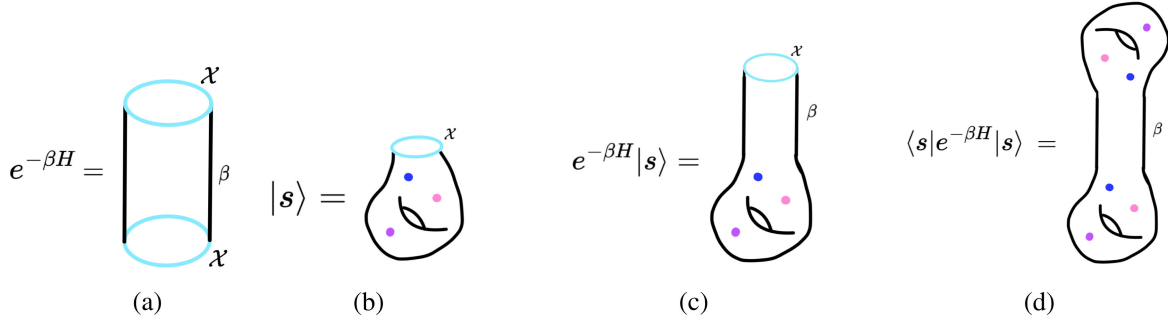


FIG. 2. Action of $e^{-\beta H}$ on states. (a) Action of the Hamiltonian, from a different perspective this is preparing the state $|2\beta\rangle$. (b) Example state in $\mathcal{H}_{\mathcal{X}}$. (c) Action of the Hamiltonian on this state. (d) Survival amplitude after Euclidean time evolution, alternatively the overlap $\langle 2\beta | s^* \rangle_L \otimes |s\rangle_R$. (adapted from [5]).

Factorization. We want to construct states in the Hilbert space of quantum gravity with two disconnected Lorentzian boundaries. To do so we slice open a Euclidean path integral so that cut on the boundary has two disconnected components $\mathcal{X}_L \cup \mathcal{X}_R$. As discussed in [6], there are two ways of constructing states in this Hilbert space. First, we could slice open the Euclidean path integral with two separate closed boundaries, one with cut $\mathcal{X}_L$ and the other with $\mathcal{X}_R$. This produces states in a tensor product Hilbert space, $\mathcal{H}_{\mathcal{X}_L} \otimes \mathcal{H}_{\mathcal{X}_R}$. Alternatively, we can slice open the path integral with a single closed connected boundary manifold with cut $\mathcal{X}_L \cup \mathcal{X}_R$. This procedure produces intrinsically two-sided states making up a Hilbert space $\mathcal{H}_{LR}$. Naively, $\mathcal{H}_{\mathcal{X}_L \cup \mathcal{X}_R} = \mathcal{H}_{LR} \cup \mathcal{H}_{\mathcal{X}_L} \otimes \mathcal{H}_{\mathcal{X}_R}$. However, the authors of [6] computed overlaps between states in $\mathcal{H}_{\mathcal{X}_L \cup \mathcal{X}_R}$ and tensor product states in $\mathcal{H}_{\mathcal{X}_L} \otimes \mathcal{H}_{\mathcal{X}_R}$ by gluing the corresponding path integrals together along the cuts.² They showed that $\mathcal{H}_{LR} = \mathcal{H}_{\mathcal{X}_L} \otimes \mathcal{H}_{\mathcal{X}_R}$, so that the complete two-boundary Hilbert space factorizes:

$$\mathcal{H}_{\mathcal{X}_L \cup \mathcal{X}_R} = \mathcal{H}_{\mathcal{X}_L} \otimes \mathcal{H}_{\mathcal{X}_R}. \quad (2.1)$$

Action of the Hamiltonian. Let H be the Hamiltonian. Then $e^{-\beta H}$ acts to glue an asymptotic boundary “cylinder” $\mathcal{X} \times \mathbb{I}_\beta$ to the cut [Figs. 2(a)–2(c)]. So the amplitude $\langle S | e^{-\beta H} | S \rangle$ is constructed by sewing the $|S\rangle$ bra and ket together via a boundary cylinder of length β [Fig. 2(d)]. The action of $e^{-\beta H}$ can also be viewed as creating a state defined on a cut consisting of two copies of $\mathcal{X}$ (think of folding the cylinder around to define a path integral sliced open to obtain a state in the two-boundary Hilbert space $\mathcal{H}_{\mathcal{X}_L \cup \mathcal{X}_R}$). We will say that $|\beta\rangle$ is a state produced in this way by the

action of $e^{-\frac{\beta H}{2}}$, so that the norm $\langle \beta | \beta \rangle$ is given by a path integral with boundary $\mathcal{X} \times S^1$ where the S^1 has period β . From this point of view, we can also think of overlaps between time evolved states in $\mathcal{H}_{\mathcal{X}}$, for example $\langle i | e^{-\frac{\beta H}{2}} | j \rangle$, as overlaps between $|\beta\rangle$ and a tensor product state, $\langle \beta | i^* \rangle_L \otimes |j\rangle_R$.

Wormholes and coarse-graining. To compute the squared overlap, $\langle i | j \rangle \langle j | i \rangle$ we insert $\mathcal{O}_{i,j}^{(\dagger)}$ on two copies of the boundary manifold and sum over bulk geometries and topologies that fill in these boundaries. This path integral has two kinds of contributions; (a) disconnected topologies coming from products of terms contributing to $\langle i | j \rangle$, and (b) connected topologies, i.e., wormholes, that interpolate between the two boundaries. The latter lead an apparent paradox; $\langle i | j \rangle = \delta_{ij}$ while $\langle i | j \rangle \langle j | i \rangle = \delta_{ij}^2 + Z_{WH}$, where Z_{WH} is the wormhole contribution. Following [5–10, 15–23] we interpret this as saying that the gravitational path integral computes coarse-grained averages over quantities in an underlying fine-grained theory. We therefore write an overbar for quantities computed from the path integral, e.g., $\overline{\langle i_1 | j_1 \rangle \langle i_2 | j_2 \rangle \cdots \langle i_n | j_n \rangle}$ denotes the path integral evaluation of the product of n overlaps.

Taking traces in an overcomplete basis. To compute the thermal partition function we must perform a trace, for which it is helpful to have a basis of states. Generic bases made by slicing the gravity path integral will not be orthogonal because of the wormhole contributions to the square of the overlap described above. Also it is often convenient in gravity to construct over-complete, and therefore nonorthogonal, bases [5, 6, 21]. Let $\{|i\rangle\}$ be such a basis with a Gram matrix of overlaps $G_{ij} \equiv \langle i | j \rangle$. Then we can construct an orthonormal basis $|v_\gamma\rangle = G_{ji}^{-1/2} U_{i\gamma} |j\rangle$; here U is a unitary transformation diagonalizing the Gram matrix ($G = U D U^\dagger$ with D diagonal), where from now on repeated indices are understood as summed (details in Appendix A of [5, 15]). Here G^{-1} is defined by analytic continuation of $(G^n)_{ij}$ from positive integer n : $G_{ij}^{-1} \equiv \lim_{n \rightarrow -1} (G^n)_{ij}$. This expression inverts nonzero eigenvalues of G and leaves zero eigenvalues untouched.

²In two-dimensional Jackiw-Teitelboim (JT) gravity, [14] showed that traces in the two-sided Hilbert space factorize into a product of two contributions to all orders in e^{-1/G_N} . Subsequently, [15] extended this result to arbitrary spacetime dimensions within the leading saddle-point approximation. A key ingredient in the arguments presented here is the identification of these two factors with the single-boundary gravitational Hilbert space, a correspondence established in [6].

$$\begin{aligned}
\overline{\text{Tr}_{\mathcal{H}_X}(e^{-\beta H})} &= \zeta \left[\text{diagram (a)} \right] + \zeta \left[\text{diagram (b)} \right] + \zeta \left[\text{diagram (c)} \right] + \dots \\
&\quad \text{(a)} \\
\overline{Z(\beta)} &= \zeta \left[\text{diagram (b)} \right] \\
&\quad \text{(b)}
\end{aligned}$$

FIG. 3. (a) The thermal partition function in gravity computed by summing over path integrals with boundaries defined by the basis of states. The colored crosses denote possible matter insertions. (b) The $\overline{Z(\beta)}$ path integral with periodic Euclidean time.

We can now write the trace as

$$\text{Tr}_{\mathcal{H}_X}(O) = G_{ij}^{-1} \langle j|O|i \rangle \equiv \lim_{n \rightarrow -1} (G^n)_{ij} \langle j|O|i \rangle, \quad (2.2)$$

and resolve the identity on $\mathcal{H}_X$ as

$$\mathbb{1}_X = \sum_{\gamma} |v_{\gamma}\rangle \langle v_{\gamma}| = G_{ij}^{-1} |i\rangle \langle j| \equiv \lim_{n \rightarrow -1} (G^n)_{ij} |i\rangle \langle j|. \quad (2.3)$$

A toolkit for fine-grained equalities. Suppose we want to use the gravitational path integral to show that $\mathcal{A} = \mathcal{B}$ in the fine-grained theory. Since the path integral only computes coarse grained averages $\bar{\mathcal{A}}$ and $\bar{\mathcal{B}}$ as discussed above, we have to show both $\bar{\mathcal{A}} - \bar{\mathcal{B}} = 0$ and $(\bar{\mathcal{A}} - \bar{\mathcal{B}})^2 = 0$ as the latter means that $\mathcal{A} - \mathcal{B} = 0$ for all the fine-grained configurations that are being averaged. In general we can only evaluate the path integral computing these quantities explicitly in the saddle-point approximation, but, as pointed out in [5], if we want to show $\bar{\mathcal{A}} = \bar{\mathcal{B}}$ it suffices to show that each contribution to the path integral $\bar{\mathcal{A}}$ is matched by an equal contribution to $\bar{\mathcal{B}}$ and vice versa. We can achieve this in three steps [5]: (1) Construct an overcomplete basis of states by inserting operators into the Euclidean path integral; (2) Take a limit in which the basis is infinitely overcomplete, leading to drastic simplifications of the topologies contributing to amplitudes; (3) Demonstrate an action-preserving bijection between geometries contributing to the path integrals of interest after this topological simplification. This task can be simplified by insertions of the identity (2.3). The authors of [5] used this procedure to demonstrate that the “two-sided shell states” of [7–9, 11, 24] form an overcomplete basis for quantum gravity with two boundaries. These methods can also be used to demonstrate a basis for quantum gravity with one boundary, and that the two-boundary Hilbert space factorizes into a product of two one-boundary Hilbert spaces [6], a fact that we will use.

A. The Gibbons-Hawking puzzle

The thermal partition function in the Hilbert space defined by the cut $\mathcal{X}$ is

$$\text{Tr}_{\mathcal{H}_X}(e^{-\beta H}) = \sum_{\gamma} \langle v_{\gamma} | e^{-\beta H} | v_{\gamma} \rangle = \sum_{\gamma} \zeta[\mathcal{M}_{\gamma}(\beta)]. \quad (2.4)$$

The right side of (2.4) is a sum over gravitational path integrals with boundary conditions appropriate to the survival amplitudes of $|v_{\gamma}\rangle$ that appear in the middle expression. We call these boundary conditions $\mathcal{M}_{\gamma}(\beta)$. By contrast, Gibbons and Hawking [4] proposed, by analogy with quantum field theory, that the thermal partition function in gravity is computed by the Euclidean path integral with periodic boundary conditions in time,

$$\overline{\text{Tr}_{\mathcal{H}_X}(e^{-\beta H})} = \overline{Z(\beta)} \equiv \zeta[\mathcal{M}_{\mathbb{S}_{\beta} \times \mathcal{X}}], \quad (2.5)$$

where $\mathbb{S}_{\beta}$ is periodic Euclidean boundary time [Fig. 3(b)]. We can also cut the periodic boundary condition into two “intervals” $\mathbb{I}_{\beta/2} \times \mathcal{X}$ of length $\beta/2$ to recognize the path integral as the norm of the state $|\beta\rangle$ defined above; $\overline{Z(\beta)} = \langle \beta | \beta \rangle$. In the Gibbons-Hawking formulation there are no asymptotic matter insertions, unlike in the states $|v_{\gamma}\rangle$ above which include both gravity and matter. This leads to a puzzle; *Why is the sum of path integrals in Fig. 3(a) equal to the Gibbons-Hawking path integral in Fig. 3(b)?*³

III. THE SINGLE-SIDED GRAVITY PARTITION FUNCTION

We now establish that the Gibbons-Hawking Euclidean gravity path integral with one periodic boundary computes a thermal trace in the Hilbert space. We will do this in two ways. First we will show that the result follows on general grounds if the two boundary Hilbert space factorizes into a product of one boundary Hilbert spaces. Second, we will explicitly perform the thermal trace in a convenient basis and show that it reduces to the Gibbons-Hawking prescription.

A. Hilbert space factorization and the thermal trace

Although the path integral $\overline{Z(\beta)}$ and sum over path integrals $\overline{\text{Tr}_{\mathcal{H}_X}(e^{-\beta H})}$ look quite different, we will use a

³See [25] for additional discussion on how to put the Gibbons-Hawking path integral on firmer footing.

resolution of the trace and identity to show that they are equal if the two boundary Hilbert space factorizes into a product of one boundary Hilbert spaces, as we have established in [6]. Consider the single-boundary gravity Hilbert space $\mathcal{H}_{\mathcal{X}}$ defined by a cut $\mathcal{X}$ with a single connected component, and let $\{|\alpha_i\rangle\}$ denote any (generally nonorthogonal and overcomplete) spanning set of states for $\mathcal{H}_{\mathcal{X}}$ made by cutting the path integral. We can resolve the trace in terms of this basis of states by using (2.2),

$$\begin{aligned}\overline{\text{Tr}_{\mathcal{H}_{\mathcal{X}}}(e^{-\beta H})} &= \lim_{n \rightarrow -1} \overline{(G^n)_{ij} \langle \alpha_j | e^{-\beta H} | \alpha_i \rangle} \\ &= \lim_{n \rightarrow -1} \overline{(G^n)_{ij} \langle \alpha_i^* | e^{-\beta H} | \alpha_j^* \rangle},\end{aligned}\quad (3.1)$$

where the last step uses the notation set up in Sec. II. Now recall that the path integral for $\overline{Z(\beta)} = \overline{\langle \beta | \beta \rangle}$ can be considered as the overlap of states corresponding to a boundary “cylinder” of length $\beta/2$. To make contact with this, we split the Hamiltonian evolution by β inside the trace into two $\beta/2$ segments and insert the resolution of the identity (2.3) on $\mathcal{H}_{\mathcal{X}}$ in between,

$$\begin{aligned}\overline{\text{Tr}_{\mathcal{H}_{\mathcal{X}}}(e^{-\beta H})} &= \overline{\text{Tr}_{\mathcal{H}_{\mathcal{X}}}(e^{-\frac{\beta H}{2}} \mathbb{1}_{\mathcal{X}} e^{-\frac{\beta H}{2}})} \\ &= \lim_{n, m \rightarrow -1} \overline{(G^n)_{ij} (G^m)_{kl} \langle \alpha_i^* | e^{-\frac{\beta H}{2}} | \alpha_k \rangle \langle \alpha_l | e^{-\frac{\beta H}{2}} | \alpha_j^* \rangle}.\end{aligned}\quad (3.2)$$

As explained in Sec. II, we can alternatively think of the operator $e^{-\beta H/2}$ as producing the two-boundary state $|\beta\rangle$, which by construction satisfies $\langle \alpha_i^* | e^{-\frac{\beta H}{2}} | \alpha_k \rangle = \langle \beta | \alpha_i \rangle_L \otimes \langle \alpha_k | \beta \rangle_R$ and similarly $\langle \alpha_l | e^{-\frac{\beta H}{2}} | \alpha_j^* \rangle = \langle \alpha_j |_L \otimes \langle \alpha_l |_R | \beta \rangle$. Using these relations we can rewrite the trace in terms of $|\beta\rangle$ and recognize two resolutions of the identity on $\mathcal{H}_{\mathcal{X}}$:

$$\begin{aligned}\overline{\text{Tr}_{\mathcal{H}_{\mathcal{X}}}(e^{-\beta H})} &= \lim_{n, m \rightarrow -1} \overline{(G^n)_{ij} (G^m)_{kl} \langle \beta | \alpha_i \rangle_L \langle \alpha_k | \beta \rangle_R \langle \alpha_j |_L \langle \alpha_l |_R | \beta \rangle} \\ &= \overline{\langle \beta | \mathbb{1}_{\mathcal{X}_L} \otimes \mathbb{1}_{\mathcal{X}_R} | \beta \rangle}.\end{aligned}\quad (3.3)$$

The state $|\beta\rangle$ is an element of the two-boundary gravity Hilbert space $\mathcal{H}_{\mathcal{X}_L \cup \mathcal{X}_R}$ (see Sec. II), while the product $\mathbb{1}_{\mathcal{X}_L} \otimes \mathbb{1}_{\mathcal{X}_R}$ is the identity on the tensor product of single-boundary theories, $\mathcal{H}_{\mathcal{X}_L} \otimes \mathcal{H}_{\mathcal{X}_R}$. As discussed in Sec. II, [6] showed that the two-boundary Hilbert space is actually a product of two single-boundary factors (2.1). Hence $\mathbb{1}_{\mathcal{X}_L} \otimes \mathbb{1}_{\mathcal{X}_R} = \mathbb{1}_{\mathcal{X}_L \cup \mathcal{X}_R}$, giving

$$\begin{aligned}\overline{\text{Tr}_{\mathcal{H}_{\mathcal{X}}}(e^{-\beta H})} &= \overline{\langle \beta | \mathbb{1}_{\mathcal{X}_L} \otimes \mathbb{1}_{\mathcal{X}_R} | \beta \rangle} = \overline{\langle \beta | \mathbb{1}_{\mathcal{X}_L \cup \mathcal{X}_R} | \beta \rangle} \\ &= \overline{\langle \beta | \beta \rangle} \equiv \overline{Z(\beta)}.\end{aligned}\quad (3.4)$$

It is easy to repeat this argument to show

$$\overline{(\text{Tr}_{\mathcal{H}_{\mathcal{X}}}(e^{-\beta H_{\mathcal{X}}}) - Z(\beta))^2} = 0, \quad (3.5)$$

since for any insertion of the trace we can make the above argument. Together (3.3) and (3.5) show that

$$\text{Tr}_{\mathcal{H}_{\mathcal{X}}}(e^{-\beta H}) = Z(\beta), \quad (3.6)$$

as a fine-grained equality.

Remarkably, the argument we have given does not require an explicit construction of the basis of states, or any particular form of the gravitational action. The argument instead relies on two requirements: (1) The gravity Hamiltonian must generate asymptotic boundary time evolution; (2) The Hilbert space of gravity with two asymptotic Lorentzian boundaries must factorize into copies of the single-boundary Hilbert space. We emphasize that, once these two criteria are granted, the chain of equalities (3.4) is purely linear algebraic; no gravitational input beyond the factorization result enters. The role of gravity is confined to establishing $\mathbb{1}_{\mathcal{X}_L} \otimes \mathbb{1}_{\mathcal{X}_R} = \mathbb{1}_{\mathcal{X}_L \cup \mathcal{X}_R}$, which is the content of [6].

Hence in any theory of quantum gravity with an (effective) path integral formulation satisfying these conditions, the Gibbons-Hawking prescription for the thermal partition function *is* in fact tracing over the states in the theory. The first requirement seems universally true in gravity theories with a timelike asymptotic boundary as it is a direct consequence of invariance under local bulk diffeomorphisms, which any gravitating system must satisfy by definition. Factorization of the Hilbert space is however nontrivial. To derive it, [6] had to include the effects of nonperturbative wormholes connecting the asymptotic boundaries used to prepare two-boundary and single-boundary gravity states. By contrast, the perturbative Hilbert space does not factorize on its own, suggesting that the Euclidean gravity path integral with periodic time does not equal the thermal trace in perturbative gravity. The sum over topologies in nonperturbative gravity is essential.

We stress at the outset that the factorization argument of Sec. III A already establishes $\text{Tr}_{\mathcal{H}_{\mathcal{X}}}(e^{-\beta H}) = Z(\beta)$ as a statement about the “full” path integral;⁴ once the two-boundary Hilbert space factorizes, the remaining manipulations are purely algebraic, so the equality holds in exactly

⁴We use “full path integral” with a caveat. In $d > 2$ the Euclidean gravitational path integral is not known to be well-defined nonperturbatively—the conformal mode renders the action unbounded below, and the sum over topologies is at best an asymptotic series—so we mean by it the formal object defined by the bulk effective theory, organized as a sum over geometries and topologies, without presuming that this object converges or admits a unique completion. The factorization argument does not require that it does; it reduces the equality to the operator identity $\mathbb{1}_{\mathcal{X}_L} \otimes \mathbb{1}_{\mathcal{X}_R} = \mathbb{1}_{\mathcal{X}_L \cup \mathcal{X}_R}$ established in [6], so the equality holds at whatever level that identity does—equivalently, in any nonperturbative completion in which the two-boundary Hilbert space factorizes—rather than relying on convergence of the sum. This is the sense intended by “full path integral equality”; it survives the inclusion of arbitrary, possibly off shell, geometries compatible with the boundary conditions (cf. Sec. 5 of [5]).

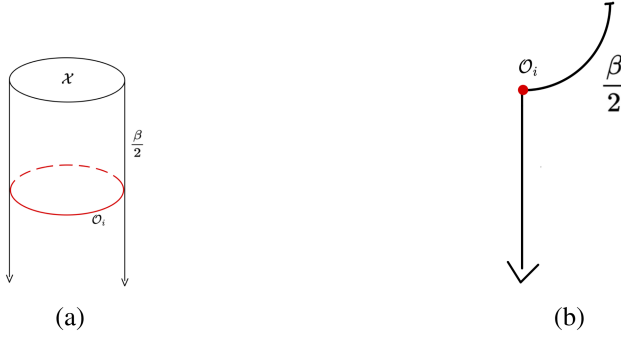


FIG. 4. Path integral boundary condition defining the single-sided shell states. (a) Euclidean boundary with topology $\mathbb{R}^{<0} \times \mathbb{S}^{d-1}$ for preparation of the shell states. The arrows indicate a half-infinite line. The shell operator $\mathcal{O}_i$ is pictured in red and $\beta/2$ is the Euclidean “preparation time”. (b) Euclidean boundary with the $\mathbb{S}^{d-1}$ suppressed. We adopt this convention, and sometimes depict this boundary with a curve or a kink to clarify diagrams (images adapted from [6]).

the same sense, and to exactly the same extent, as the factorization identity itself, with no reference to a saddle expansion. The explicit evaluation in this section is therefore best read as an independent and more concrete check of that conclusion, one that operates mostly at the level of the sum over saddles.

B. Wormholes, a basis, and the trace

As an alternative approach, we will explicitly perform the thermal trace in the “shell state” basis for the single boundary Hilbert space [6,13]. To define these states we fix a boundary with topology $\mathbb{R}^{<0} \times \mathbb{S}^{d-1}$ (where $\mathbb{R}^{<0}$ is the half line) and insert an $\mathbb{S}^{d-1}$ symmetric heavy [$m \sim O(1/G_N)$] dust shell operator $\mathcal{O}_S$ separated by a boundary length $\frac{\beta}{2}$ from the cut $\mathcal{X}$. Thus, the boundary time extends from $-\infty$ to $\tau = 0$, where $\mathcal{O}_S$ is inserted, followed by a time section up to $\tau = \beta/2$ (Fig. 4). We can obtain an infinite family of shell states in this way by varying the mass m_i of the operators $\mathcal{O}_i$ to produce states $|i\rangle$.

We will need to consider path integrals with a boundary consisting of a Euclidean half line $[-\infty, 0]$ followed by operator insertions and Euclidean time evolution up to time T_E , and then another half line $[T_E, \infty]$.⁵ We will call T_E the “length of the strip” and denote the path integral with this boundary condition as $S(T_E)$. For example, to compute the Gram matrix element $G_{ij} \equiv \langle i|j \rangle$ we construct a ket by the path integral in Fig. 4, and a bra defined by a similar path integral on $\frac{\beta}{2} < \tau < \infty$. The boundary condition defining G_{ij} is thus an interval $[-\infty, 0]$ followed by $\mathcal{O}_j$ insertion, boundary time evolution by an amount β , $\mathcal{O}_i^\dagger$

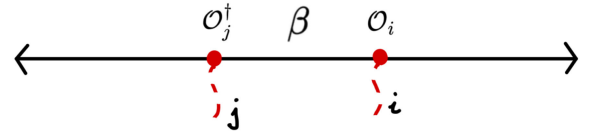


FIG. 5. Shell-strip asymptotic boundary condition for the overlap $\langle j|i \rangle$ consisting of the line $\lim_{\alpha \rightarrow \infty} [-\alpha, \beta + \alpha]$ on which $\mathcal{O}_i$ and $\mathcal{O}_j^\dagger$ are inserted at $\tau = 0$ and $\tau = \beta$, respectively (images adapted from [6]).

insertion, and another half line $[\beta, \infty]$ (Fig. 5). In the above terminology this is a strip of length β containing $\mathcal{O}_j$ and $\mathcal{O}_i^\dagger$ operator insertions. The coarse-grained overlap $\langle i|j \rangle$ is evaluated by the path integral with such a *shell-strip boundary condition* (Fig. 6). Following [8], we make different shell states orthogonal, $\langle i|j \rangle = \delta_{ij} Z_1$, by taking their inertial mass differences to be arbitrarily large, since it will take $|m_i - m_j|$ bulk interactions in Planck units for such shells to annihilate.

Norm. As shown in [6], we can construct saddle points for path integrals with the shell strip boundary condition by filling in one side of the shell world volume with any saddle geometry of $\overline{Z(\beta)} = \langle \beta|\beta \rangle$ (the “disk”) and the other with a Euclidean vacuum geometry with noncompact time (the “strip”). We use the Israel junction conditions [26] to glue the geometries at the shells (Fig. 6). For example, suppose we take a Euclidean anti-de Sitter (AdS) strip with a shell operator $\mathcal{O}_i$ on the boundary at $\tau = 0$ and the conjugate operator $\mathcal{O}_i^\dagger$ at time $\tau = \Delta T_S$ and a Euclidean AdS disk with boundary insertions $\mathcal{O}_i$ and $\mathcal{O}_i^\dagger$ separated by ΔT_D on a disk of boundary circumference $\beta + \Delta T_D$. To construct the saddle-point geometry we discard the shell homology regions (purple in Fig. 6) from the strip and disk and use the junction conditions to glue the shell world volumes.

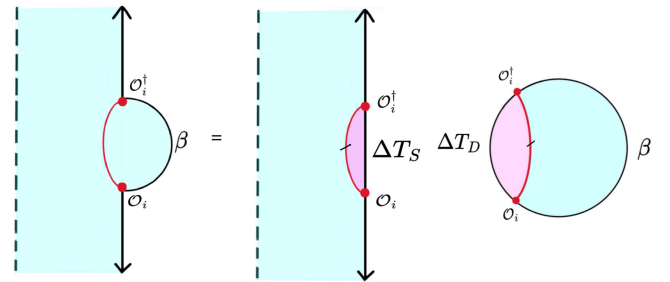


FIG. 6. The saddle for the norm $\langle i|i \rangle$ is constructed by considering the shell propagating on a disk and strip separately for some propagation times $\Delta T_{S,D}$ and then gluing them together along the i -shell world volume by discarding the shell homology region (purple). The junction conditions dynamically determine $\Delta T_{S,D}$ to yield an on shell glued geometry. We have suppressed the angular directions in these diagrams, and represent the radius-time plane with the dashed line representing the origin (images adapted from [6]).

⁵See [6] and Fig. 5 for a specification of the limit procedure in which these strip boundary conditions are defined. We do not specify the details here as we will not need them.

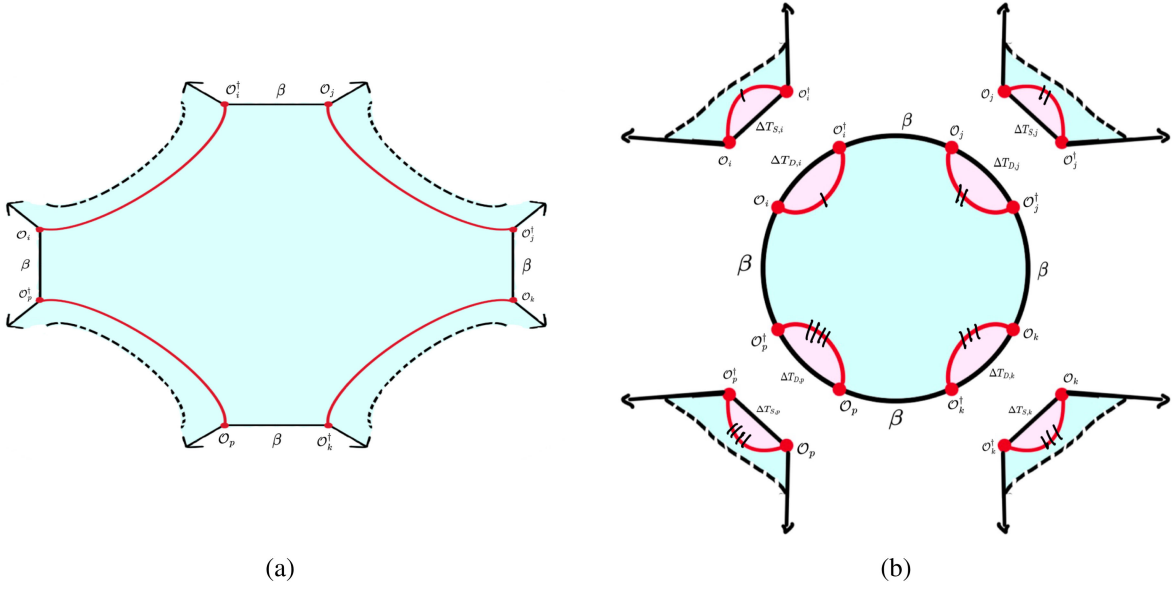


FIG. 7. The fully connected single-sided shell wormhole saddles for $\overline{(G^n)_{ii}}$ are constructed by gluing n strips into the disk along the shell world volumes, depicted here for $n = 4$. Images adapted from [6].

The junction conditions dynamically determine ΔT_S and ΔT_D such that the resulting geometry satisfies the equations of motion.

Large shell mass limit. As discussed in [6], as $m_i \rightarrow \infty$ limit the turning point of the shell approaches the asymptotic boundary and $\Delta T_S, \Delta T_D \rightarrow 0$. So the shell homology regions pinch off and the remainder grows to cover the entire strip and disk. In this limit, the shell contribution to the action is a universal factor Z_{m_i} and the overlap is

$$\overline{\langle i|j \rangle} = \delta_{ij} Z_{m_i} \times \bar{Z}(\beta) \times \bar{S}(0). \quad (3.7)$$

As the above construction could have been done using any of the saddles for $\bar{Z}(\beta)$, (3.7) equates the sum over saddle points of the lhs to that of the rhs.

Wormholes. To construct fully connected wormhole saddles contributing to

$$\overline{(G^n)_{ii}} = \overline{\langle i|j_1 \rangle \langle j_1|j_2 \rangle \cdots \langle j_n|i \rangle} \quad (3.8)$$

(no sum on i, j_k) we connect n boundaries [from each factor in $(G^n)_{ii}$] through a single wormhole. The k th shell must propagate from $\mathcal{O}_k$ to $\mathcal{O}_k^\dagger$ and so the wormhole connectivity reflects the index structure of $\overline{(G^n)_{ii}}$ [Fig. 7(a)]. Similarly to the norm, we construct the saddle by gluing n copies of the Euclidean strip into a disk along the shell world volumes [Fig. 7(b)]. To construct saddle geometries satisfying this boundary condition consider a disk with n shells inserted on the asymptotic boundary, $\mathcal{O}_{S_i}$ separated from $\mathcal{O}_{S_i}^\dagger$ by boundary time $\Delta T_{D,i}$, and $\mathcal{O}_{S_i}^\dagger$ from $\mathcal{O}_{S_{i+1}}$ by boundary time β in a circular pattern. Separate shell strips are constructed

by placing a shell $\mathcal{O}_{S_i}$ on the strip boundary separated from $\mathcal{O}_{S_i}^\dagger$ by $\Delta T_{S,i}$ boundary time. We then remove the shell homology regions on the disk and strips and use the junction conditions to identify corresponding shell world volumes, and to dynamically determine $\Delta T_{D,i}$ and $\Delta T_{S,i}$. Details are in [6]. As described above, the shell propagation times go to zero in the large shell-mass limit, so the saddle becomes a product of n strips, a disk of length $n\beta$, and universal shell contributions; $\overline{(G^n)_{ii}} = \bar{Z}(n\beta) \times \bar{S}(0)^n \times \prod_{i=1}^n Z_{m_i}$. Normalizing shell states to have unit norm we get

$$|k\rangle \rightarrow \frac{|k\rangle}{\sqrt{Z_{m_k} \times \bar{Z}(\beta) \times \bar{S}(0)}}; \quad \overline{(G^n)_{ii}} = \frac{\bar{Z}(n\beta)}{\bar{Z}(\beta)^n}. \quad (3.9)$$

Saddle geometries. Following [6], to associate geometry to a shell state we consider the leading gravitational saddle point for the norm $\overline{\langle i|i \rangle}$. In the large shell mass limit resulting in (3.7), this is the determined by the leading saddle for $\bar{Z}(\beta)$. In the asymptotically AdS case the leading saddle is either thermal Euclidean AdS if $\beta > \beta_{HP}$ or the large Euclidean black hole if $\beta < \beta_{HP}$, where β_{HP} marks the Hawking-Page transition [27]. With asymptotically flat boundary conditions, thermal flat space is always leading, although after a microcanonical projection the black hole saddle can be leading. The time reflection symmetric slice of these saddles is continued to Lorentzian signature. There are thus two classes of single-sided shell states:

- (i) *Type A:* The leading saddle is the Euclidean black hole, so the Lorentzian geometry is a black hole with a shell in the interior, capped off by the vacuum

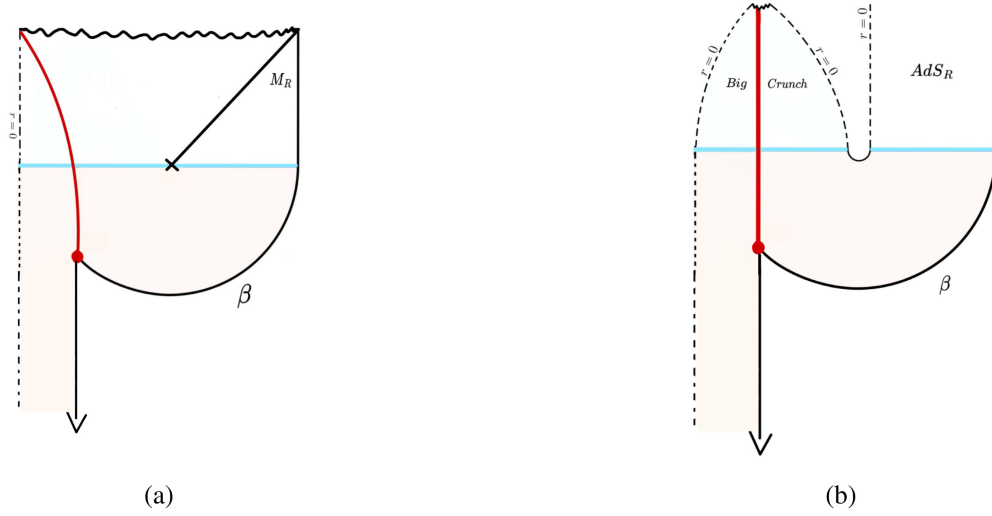


FIG. 8. Analytic continuation of the single-sided shell state saddle points to Lorentzian signature, shown here with asymptotically AdS boundary conditions. For asymptotically flat boundary conditions the vertical lines are replaced by diamonds. (a) Type A shell state corresponding to a single-sided black hole. (b) Type B shell state consisting of thermal AdS (or thermal flat space) with and added disconnected compact Big Crunch AdS cosmology. Images adapted from [6].

behind the shell. We will refer to these as *Type A* states. In AdS space we find these saddles when $\beta < \beta_{HP}$ and in asymptotically flat space in the finite energy microcanonical ensemble;

- (ii) *Type B*: The leading saddle has a noncontractible thermal circle and the Lorentzian geometry is the thermal space with a disconnected closed “Big Crunch” universe in which the shell propagates. We will refer to these states as *Type B* shell states. In AdS space we find these saddles for $\beta > \beta_{HP}$ and in asymptotically flat space for any β in the canonical ensemble.

1. Explicit trace

Let $\mathcal{H}_\chi$ be the single boundary Hilbert space. To establish the fine-grained equality $\mathcal{A} = \text{Tr}_{\mathcal{H}_\chi}(e^{-\beta H_\chi}) = Z(\beta) = \mathcal{B}$, we will use the toolkit from [5] (see Sec. II) to show that $\overline{\mathcal{A} - \mathcal{B}} = 0$ and $\overline{(\mathcal{A} - \mathcal{B})^2} = 0$ directly in the single boundary theory. Denote a single-sided shell state of mass m_i and preparation temperature β_R as $|i\rangle_R$, and the Hilbert space spanned by κ_R such states of varying mass but fixed preparation temperature as $\mathcal{H}_R$. The single-sided shell states span $\mathcal{H}_\chi$ as $\kappa_R \rightarrow \infty$ [6] (we return below to the weaker condition $\kappa_R \gtrsim e^S$ that suffices in the microcanonical ensemble). Hence, the trace over $\mathcal{H}_\chi$ can be resolved as

$$\overline{\text{Tr}_{\mathcal{H}_\chi}(e^{-\beta H_\chi})} = \overline{\text{Tr}_{\mathcal{H}_R}(e^{-\beta H_\chi})} = \lim_{n \rightarrow -1} \overline{(G^n)_{ij} \langle j | e^{-\beta H_\chi} | i \rangle}, \quad (3.10)$$

where we are summing over i, j .

Consider the saddle-point approximation of $\overline{(G^n)_{ij} \langle j | e^{-\beta H_\chi} | i \rangle}$. As discussed above, powers of the

single-sided Gram matrix contribute shell-strip insertions of length β_R (Fig. 5), while $\langle j | e^{-\beta H_\chi} | i \rangle$ contributes to a lengthened shell strip boundary of length $\beta + \beta_R$. To evaluate the gravitational path integral we should sum over topologies. However, in the $\kappa_R \rightarrow \infty$ limit in which the shell states span $\mathcal{H}_\chi$, only geometries that do not break any shell index loops contribute [5]. For (3.10) this means that contributing geometries fill in the boundary index loop [Fig. 9(a)] with a single fully-connected bulk. This loop structure of the trace will lead to the circular boundary condition defining $Z(\beta)$.

These fully-connected saddle geometries are the same as those constructed for (3.8) above, except that the $\mathcal{O}_{S_{n+1}}^\dagger$ and $\mathcal{O}_{S_1}$ shell insertions on the boundary are separated by $\beta + \beta_R$ because of the additional $e^{-\beta H_\chi}$ time evolution. So, by the same logic, in the large shell mass limit the trace equals $\bar{Z}(\beta + (n+1)\beta_R) \times \bar{S}(0)^{n+1} \times \prod_{i=1}^{n+1} Z_{m_i}$, and after normalizing the shell states as in (3.9) yields

$$\overline{(G^n)_{ij} \langle j | e^{-H_R \beta} | i \rangle} = \kappa_R^{n+1} \frac{\bar{Z}(\beta + (n+1)\beta_R)}{Z(\beta_R)^{n+1}}. \quad (3.11)$$

Hence, in the $n \rightarrow -1$ limit,

$$\overline{\text{Tr}_{\mathcal{H}_\chi}(e^{-\beta H_\chi})} = \overline{Z(\beta)}. \quad (3.12)$$

This equality includes all the saddle points contributing to the left and right sides.

As explained in [5], in the $\kappa_R \rightarrow \infty$ limit we can readily repeat this argument to also show $\overline{(\text{Tr}_{\mathcal{H}_\chi}(e^{-\beta H_\chi}) - Z(\beta))^2} = 0$ establishing the fine-grained equality $\text{Tr}_{\mathcal{H}_\chi}(e^{-\beta H}) = Z(\beta)$ in the saddle-point

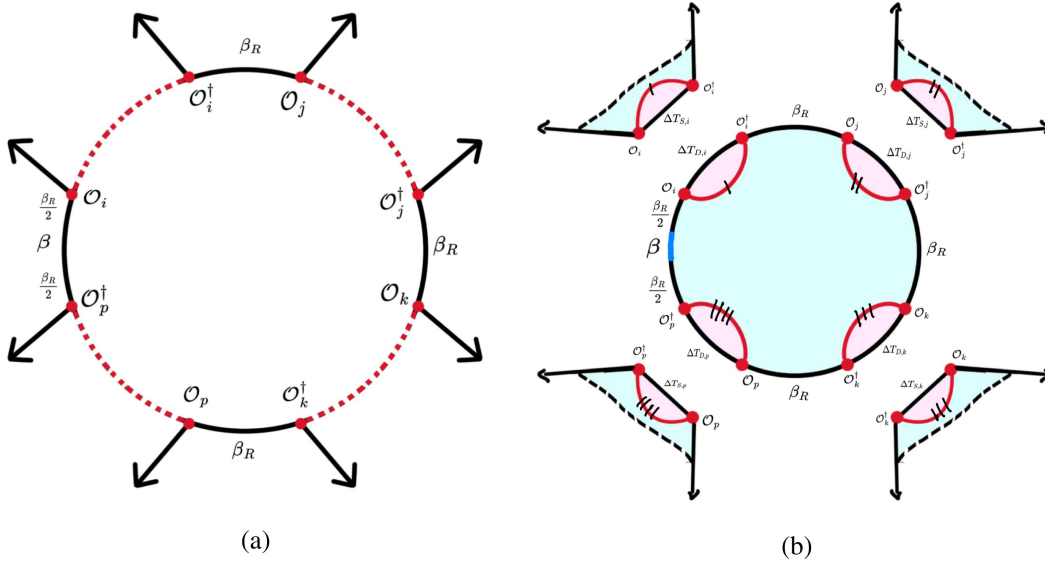


FIG. 9. (a) Boundary condition for the path integral computing (3.10) with $n = 3$. (b) Construction of the semiclassical saddle for (3.10) by gluing shell strips into a disk.

approximation. Once again, the relation (3.12) is only true in the nonperturbative theory, and would fail in a perturbative description that did not include wormhole contributions to the path integral.

Matching the full path integral. The equality (3.12) extends from the leading saddles to the full path integral. The cleanest statement of this is the factorization argument of Sec. III A; there, $\text{Tr}_{\mathcal{H}_X}(e^{-\beta H}) = \langle \beta | \mathbb{1}_{\mathcal{H}_L} \otimes \mathbb{1}_{\mathcal{H}_R} | \beta \rangle = \langle \beta | \beta \rangle = Z(\beta)$ followed purely algebraically once the two-boundary Hilbert space factorizes, and is therefore valid for the full path integral whenever the latter is suitably defined, independently of any saddle expansion. The explicit shell-basis computation of this section provides an independent check of the same statement, largely at the level of the sum over saddles. It can be extended beyond the leading saddle by a heuristic argument, which we sketch but do not aim to make precise, as follows. First, the saddle geometries of the two sides are in one-to-one correspondence. The shell strip admits a single saddle, while the central disk into which the strips are glued admits one saddle for each saddle of the Gibbons-Hawking path integral $\overline{Z(\beta)}$ —these disk geometries are precisely the $\overline{Z(\beta)}$ saddles being glued in. Hence, every saddle contributing to the explicit trace corresponds to exactly one saddle of $\overline{Z(\beta)}$, and vice versa. Second, the perturbative corrections around each pair of corresponding saddles match. In the large shell-mass limit each strip decouples into an independent region carrying its own perturbative expansion (one may, for instance, dress each strip with its own loop corrections), and the heavy shells localize onto the asymptotic boundary and are classical, so they contribute no fluctuations of their own. With standard (Dirichlet) asymptotically locally AdS boundary conditions the leading

boundary data is held fixed, and the bulk fluctuations the shells could couple to are normalizable modes that fall off towards the boundary; their contribution in the thin near-boundary region onto which the shells localize is subleading in the large-mass limit, so to the order we work the perturbative corrections reside in the central disk, where they coincide with those of $\overline{Z(\beta)}$ around the corresponding saddle. The strip and shell contributions cancel against the unit-norm normalization (3.9). In this sense the bulk loop expansions around corresponding saddles agree order by order, and the explicit trace reproduces $\overline{Z(\beta)}$ beyond the leading saddle; the same correspondence applied to $\overline{(\text{Tr}_{\mathcal{H}_X}(e^{-\beta H}) - Z(\beta))^2}$ gives $\overline{(\text{Tr}_{\mathcal{H}_X}(e^{-\beta H}) - Z(\beta))^2} = 0$. We do not attempt to make this matching precise; it rests on the suppression of the near-boundary fluctuations coupling to the heavy shells in the large-mass limit, a statement we expect but do not establish here. We therefore rely on the factorization argument for the clean full-path-integral equality.

On the large κ limit. The role of the $\kappa_R \rightarrow \infty$ limit differs depending on the ensemble. In the canonical ensemble the single-boundary Hilbert space is infinite dimensional, so $\kappa_R \rightarrow \infty$ is genuinely required for the shell states to span; in that limit every geometry breaking a shell index loop is suppressed, leaving no surviving $O(1/\kappa_R)$ corrections. In the microcanonical ensemble at an energy above the black-hole threshold the Hilbert space is finite dimensional, of dimension $\sim e^S$ with S the Bekenstein-Hawking entropy, and the shell states already span once $\kappa_R \gtrsim e^S$, as shown in [6,8,9]. For such finite κ_R the equality is recovered in a planar limit, $\kappa_R, e^{1/G_N} \rightarrow \infty$ with $\kappa_R e^{-1/G_N} \sim \mathcal{O}(1)$ held fixed, in which a resolvent resummation (see also [10,21]) incorporates the additional topologies; we do not carry out

this resummation explicitly, as the methods are standard and the result not especially illuminating. In either case $\text{Tr}_{\mathcal{H}_\chi}(e^{-\beta H}) = Z(\beta)$ holds because one is tracing over a genuine basis—shown in [6] for the single boundary case, to exist microcanonically once $\kappa_R \gtrsim e^S$ and canonically as $\kappa_R \rightarrow \infty$.

C. What is the Bekenstein-Hawking entropy counting?

In thermal physics, the microcanonical density of states is obtained from the partition function by a Laplace transform $\rho(E) = \int d\beta e^{\beta E} Z(\beta)$. This density of states is often expressed as the exponential of the microcanonical entropy. Thus, given (3.12), we should be able to read off the microcanonical entropy in gravity in the saddle-point approximation by computing the Laplace transform of $\bar{Z}(\beta)$. The path integral for $\bar{Z}(\beta)$ can have multiple saddle points. The value of the entropy, and its interpretation in terms of microstates, depend on which saddle point dominates. For example, in asymptotically AdS gravity $\bar{Z}(\beta)$ has three saddle geometries; the large black hole, thermal AdS and the small black hole. The small black hole is always subdominant, while the Hawking-Page transition determines dominance between the other two [27]. Above the transition, the Laplace transform of $\bar{Z}(\beta)$ is dominated by the black hole saddle, and the explicit saddle point computation yields the Bekenstein-Hawking entropy formula $S = A/4G$ where A is the area of the horizon in the black hole of energy (mass) E . Similar computations lead to $S = A/4G$ in any theory which is described by Einstein gravity at low energies regardless of the cosmological constant. This is why the Bekenstein-Hawking entropy formula is usually described as being universal.

What are the states counted by the Bekenstein-Hawking entropy formula? Since the formula arises from a dominant saddle point of the Gibbons-Hawking path integral which has a black hole geometry, we expect that states contributing to the entropy can be regarded as black hole microstates. Any such microstate should have a black hole exterior geometry, but could have structure behind the horizon that cannot be measured by the asymptotic observer. However, our derivation in Sec. III A and Sec. III B just required a complete basis of states, and we could have used the Type B shell states, which do not contain horizons [Fig. 8(b)]. So what are the states counted by the Bekenstein-Hawking entropy formula, and can they be interpreted as black hole microstates?

To resolve this question, we have to examine a microcanonical fixed energy Hilbert space $\mathcal{H}_\chi|_E$. We denote the projector into the microcanonical Hilbert space $\mathcal{H}_\chi|_E$ as $\prod_E$. Let us denote a single-sided shell state formed by insertion of the $\mathcal{O}_i$ shell operator and time evolution by $\frac{\beta_R}{2}$ as $|i^{\beta_R}\rangle$. Isolating the $\frac{\beta_R}{2}$ time evolution we write this as $|i^{\beta_R}\rangle = e^{-\frac{\beta_R}{2}H}|\hat{i}\rangle$. Labeling the energy eigenstates of H as

$\{|n\rangle\}$ we write the shell states in the energy basis as $|i^{\beta_R}\rangle = \sum_n e^{-\frac{\beta_R}{2}E_n} \langle n|\hat{i}\rangle |n\rangle$ and the microcanonically projected states as $\prod_E |i^{\beta_R}\rangle = \sum_{E_n=E} e^{-\frac{\beta_R}{2}E} \langle n|\hat{i}\rangle |n\rangle$. States with the same $\mathcal{O}_i$ insertion but different values of $\frac{\beta_R}{2}$ therefore result in linearly dependent microcanonical states, differing only by a factor of $e^{-\frac{\beta_R}{2}E}$. Ignoring this prefactor, the projection can be performed by the Laplace transform $\prod_E |i\rangle \equiv |i^E\rangle \sim \int d\beta e^{\beta E/2} |i^\beta\rangle$.

Any state $|\Psi_E\rangle \in \mathcal{H}_\chi|_E$ can be written as a linear combination $|\Psi_E\rangle = \sum_i \alpha_i |i\rangle$ of the canonical shell states, as they span the full Hilbert space $\mathcal{H}_\chi$. We must have $\prod_E |\Psi_E\rangle = |\Psi_E\rangle$ and therefore any state $|\Psi_E\rangle \in \mathcal{H}_\chi|_E$ can be written as $|\Psi_E\rangle = \sum_i \alpha_i \prod_E |i\rangle \equiv \sum_i \alpha_i |i^E\rangle$. Hence, the projected shell states $|i^E\rangle$ span $\mathcal{H}_\chi|_E$ because the canonical states $|i\rangle$ span the full Hilbert space $\mathcal{H}_\chi$.

The semiclassical geometry associated to the microcanonical shell states is that of the leading saddle contribution to its norm $\overline{\langle i^E|i^E\rangle} = \overline{\langle i|\prod_E|i\rangle} \sim \int d\beta e^{\beta E} \langle i^\beta|i^\beta\rangle$, where we have prepared the canonical state $|i\rangle$ at some arbitrary inverse temperature $\frac{\beta}{2}$. As discussed in Sec. III B there are two possible leading saddles to the canonical overlap; the Type A and Type B saddle. The Laplace transform will pick out the Type A saddle at microcanonical energies above the black hole threshold. To see this note the canonical overlap is proportional to the Gibbons-Hawking path integral $\overline{\langle i^\beta|i^\beta\rangle} = Z_{m_i} \bar{S}(0) \bar{Z}(\beta + \tilde{\beta})$ (see Sec. III B). We have shown that $\bar{Z}(\beta)$ computes the thermal partition function of the gravity theory, and hence $\int d\beta e^{\beta E} \bar{Z}(\beta + \tilde{\beta}) = e^{-\tilde{\beta}E + S(E)}$ where $S(E)$ is the microcanonical entropy. Hence, $\overline{\langle i^E|i^E\rangle} = \bar{S}(0) Z_{m_i} e^{-\tilde{\beta}E + S(E)}$. As shown above, for energies above the black hole threshold (if there is one) the microcanonical entropy is given by the Bekenstein-Hawking formula and therefore the Laplace transform must have picked out the black hole saddle. Hence the geometry associated to the microcanonical shell state is that of a Type A shell state; that is a single-boundary black hole geometry [Fig. 8(a)] with a shell behind the horizon, and an exterior identical to that of the black hole saddle to $\bar{Z}(\beta)$.

Thus, we see that at energies above the black hole threshold, the basis for the micro-canonical Hilbert space consists solely of black hole microstates, or, more precisely, states for which the leading saddle computing the norm is a black hole microstate.⁶ Furthermore, by similar logic the microcanonical entropy of the two-boundary theory is given by the sum of the Bekenstein-Hawking entropies of the two horizons, $S_{E_L} + S_{E_R}$, and the states that are being counted are microstates of the two-boundary black hole in

⁶It should be possible to carry out a similar analysis in the context of single-sided states with end-of-the-world branes behind the horizon [28], although these have not been shown to span the full Hilbert space.

the sense discussed above. This follows from the two-boundary shell state construction in [8–11,15] which was shown to provide a basis in [5].

IV. ENTANGLEMENT ENTROPY AND TRACES IN THE GRAVITY HILBERT SPACE

Let us consider two entangled universes, namely states of a gravitating system with two disconnected Lorentzian boundaries. Examples include the thermofield double [29] and PETS states [30] in asymptotically AdS spacetime, and the two-sided black hole microstates in [5,8,10,11].⁷ The full two-boundary Hilbert space $\mathcal{H}_{\mathcal{X}_L \cup \mathcal{X}_R}$ factorizes into copies of the single-boundary theories $\mathcal{H}_{\mathcal{X}_L} \otimes \mathcal{H}_{\mathcal{X}_R}$ [6], and the generic state in this Hilbert space will have some nonzero entanglement between these factors. Given such a

state, $|\Psi\rangle \in \mathcal{H}_{\mathcal{X}_L \cup \mathcal{X}_R}$, we would like to compute the n th Rényi entropy,

$$\overline{\text{Tr}_{\mathcal{H}_{\mathcal{X}_R}}(\rho_R^n)}, \quad (4.1)$$

where $\rho_R = \text{Tr}_{\mathcal{H}_{\mathcal{X}_L}}(|\Psi\rangle\langle\Psi|)$ is the reduced density matrix on R .⁸

As discussed above, $\mathcal{H}_{\mathcal{X}_{L,R}}$ are spanned by the single-sided shell states [6]. We can therefore use the shell state resolution of the trace (2.2) to explicitly trace out one of the universes to obtain the reduced density matrix on the other. For example, tracing out $\mathcal{H}_{\mathcal{X}_L}$ results in $\rho_R = \lim_{m \rightarrow -1} (G^m)_{L,ab} \langle b|_L \Psi \rangle \langle \Psi|a \rangle_L$ and $\rho_R^n = \lim_{p_1, \dots, p_n \rightarrow -1} \prod_{i=1}^n (G^{p_i})_{L,a_i b_i} \langle b_i|_L \Psi \rangle \langle \Psi|a_i \rangle_L$. Performing the final trace over $\mathcal{H}_{\mathcal{X}_R}$ we obtain:

$$\overline{\text{Tr}_{\mathcal{H}_{\mathcal{X}_R}}(\rho_R^n)} = \lim_{k, p_1, \dots, p_n \rightarrow -1} (G^k)_{R,cd} \langle d|_R \prod_{i=1}^n (G^{p_i})_{L,a_i b_i} \langle b_i|_L \Psi \rangle \langle \Psi|a_i \rangle_L |c \rangle_R. \quad (4.2)$$

For concreteness, and without loss of generality, we will consider states of the form $|\Psi(\tilde{\beta}_L, \tilde{\beta}_R)\rangle = |e^{-\frac{\tilde{\beta}_L H_L}{2}} \mathcal{O}_\Psi e^{-\frac{\tilde{\beta}_R H_R}{2}}\rangle$, where $\mathcal{O}_\Psi$ can be nonlocal and may involve time evolution. The boundary conditions defining the density matrix ρ_R^n and the (unnormalized) Rényi entropy (4.2) are depicted in Figs. 10(a) and 10(b). We can evaluate the path integral subject to these boundary conditions explicitly within the saddle-point approximation, and then take the $k, p_1, \dots, p_n \rightarrow -1$ limit to obtain the Rényi entropy.

In holographic theories of gravity the boundary replica trick motivates a different boundary condition for the path integral computing Rényi entropy. In particular, the holographic boundary condition for $\text{Tr}_{\mathcal{H}_{\mathcal{X}_R}}(\rho_R^n)$, denoted $Z_\circ(\rho_\Psi^n)(\tilde{\beta}_L, \tilde{\beta}_R)$, corresponds to a “torus” formed by n alternating $\mathcal{O}_\Psi, \mathcal{O}_\Psi^\dagger$ insertions, with $\mathcal{O}_{\Psi,i}, \mathcal{O}_{\Psi,i}^\dagger$ separated by $\tilde{\beta}_L$ and $\mathcal{O}_{\Psi,i}^\dagger, \mathcal{O}_{\Psi,i+1}$ separated by $\tilde{\beta}_R$ boundary time (see Fig. 12). Indeed, in AdS/CFT this is precisely the prescription that leads to the bulk derivation [34] of the Ryu-Takayanagi and Hubeny-Rangamani-Takayanagi formulas [35–37] relating the entanglement entropy of a boundary subregion in the conformal field theory to area of an extremal surface in the bulk homologous to that subregion. This leads to a puzzle analogous to one we discussed in Sec. III for the Gibbons-Hawking path integral; *Why does the holographic replica path integral prescription for the Rényi entropy [Fig. 12(c)] reproduce the trace over the Hilbert space computed by the sum over path integrals with boundary conditions in Fig. 10?*

Concretely, we must show that the gravity Hilbert space prescription and the holographic boundary replica prescription for computing the Rényi entropy are the same at the fine-grained level:

$$\text{Tr}_{\mathcal{H}_{\mathcal{X}_R}}(\rho_R^n) \stackrel{?}{=} Z_\circ(\rho_\Psi^n)(\tilde{\beta}_L, \tilde{\beta}_R). \quad (4.3)$$

Comparing Fig. 10 and Fig. 11 with Fig. 12 we want to show from the gravity Hilbert space perspective that: (1) the trace over $\mathcal{H}_{\mathcal{X}_L}$ in the definition of the reduced density matrix ρ_R sews the Ψ bra and ket together along the $\mathcal{X}_L$ cut to produce a “cylinder,” and (2) the trace over $\mathcal{H}_{\mathcal{X}_R}$ in $\text{Tr}_{\mathcal{H}_{\mathcal{X}_R}}(\rho_R^n)$ sews copies of the cylinder together into the “torus” of Fig. 12(c). Analogously to the discussion in Sec. II A these two requirements are not immediate from the rules for cutting the gravitational path integral to obtain states that were outlined at the beginning of Sec. II.

We will provide two arguments showing (4.3). The first uses strategic insertions of the identity to show (1) is true in general, and show requirement (2) from factorization of the two-boundary Hilbert space. Second, we evaluate the traces in (4.2) explicitly in the single-sided shell basis and find equality between the two prescriptions.

A. The trace glues density matrices

Consider the boundary condition defining the reduced density matrix [Fig. 10(a)],

⁷Entanglement can also be introduced between de Sitter universes which not have a Euclidean boundary, and asymptotically AdS or flat universes that do [31–33], but we will not examine those cases here.

⁸Strictly speaking the Rényi entropy is $\log[\overline{\text{Tr}_{\mathcal{H}_{\mathcal{X}_R}}(\rho_R^n)}] - \log[(\text{Tr}_{\mathcal{H}_{\mathcal{X}_R}}(\rho_R))^n]$, but we will work with the simpler quantity (4.1) from which the entropy can be easily calculated, and abuse terminology to call it Rényi entropy.

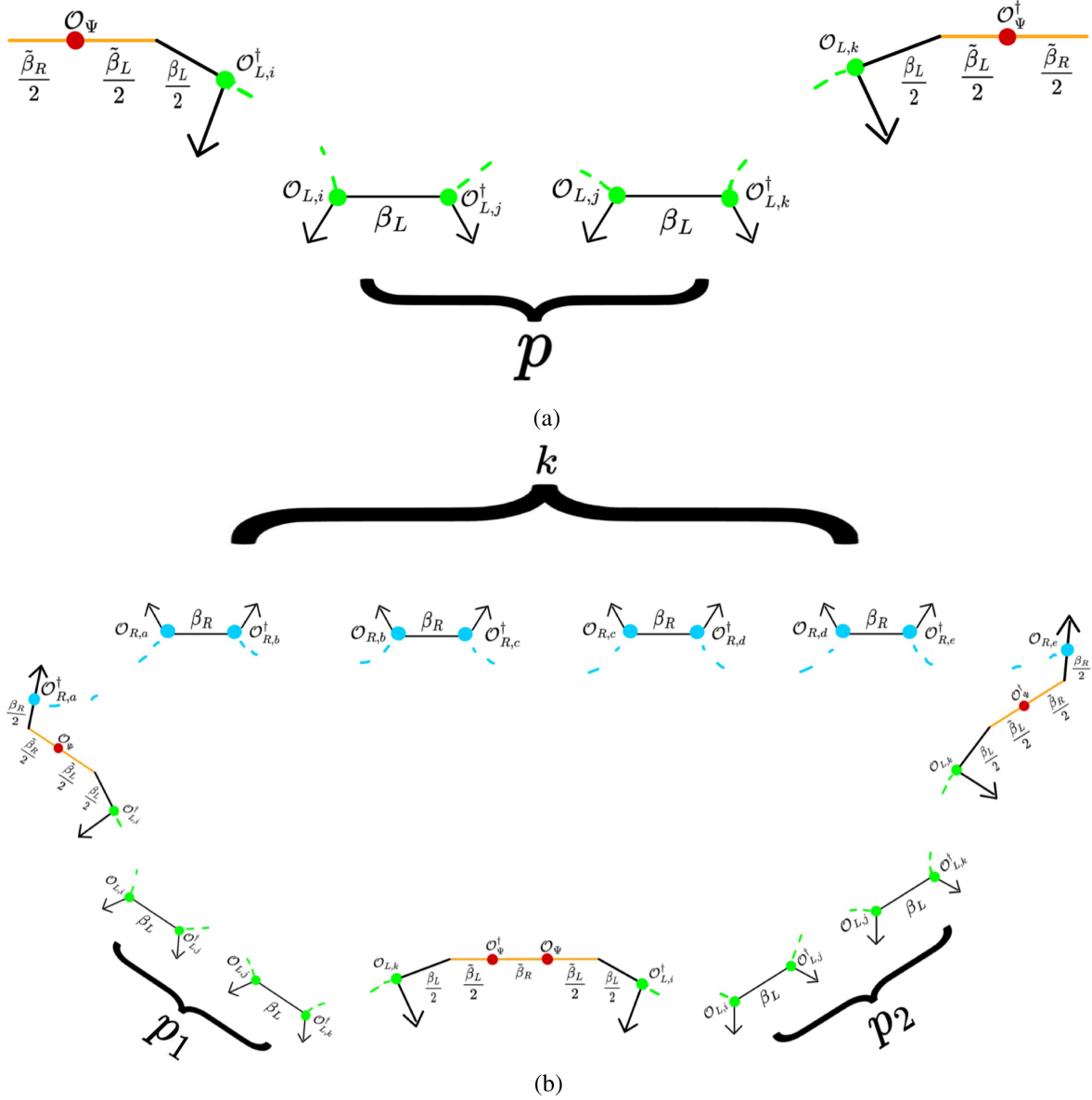


FIG. 10. (a) Boundary condition for $\rho_R = \text{Tr}_{\mathcal{H}_{\mathcal{X}_L}}(|\Psi\rangle\langle\Psi|)$. (b) Boundary condition defining $\text{Tr}_R(\rho_R^n)$, depicted for $n = 2$.

$$\rho_R = \text{Tr}_{\mathcal{H}_{\mathcal{X}_L}}(|\Psi\rangle\langle\Psi|) = \lim_{m \rightarrow -1} (G^m)_{L,ab} \langle b|_L \Psi \rangle \langle \Psi|a\rangle_L. \quad (4.4)$$

We will argue that the role of the trace over the left Hilbert space is to sew the L parts of the $|\Psi\rangle$ bra and ket together into a boundary condition given by $\frac{\tilde{\beta}_R}{2}$ time evolution, O_Ψ insertion, $\tilde{\beta}_L$ time evolution, $O_\Psi^\dagger$ insertion and finally another $\frac{\tilde{\beta}_R}{2}$ time evolution [Fig. 12(b)]. To see this, note that the state $|\Psi\rangle$ can alternatively be considered as a map $M_\Psi: \mathcal{H}_{\mathcal{X}_L} \rightarrow \mathcal{H}_{\mathcal{X}_R}$ that starts with a state in $\mathcal{H}_{\mathcal{X}_L}$, time evolves by $\frac{\tilde{\beta}_L}{2}$, applies to operator O_Ψ and then time evolves by $\frac{\tilde{\beta}_R}{2}$ up to the cut $\mathcal{X}_R$ to define a state in $\mathcal{H}_{\mathcal{X}_R}$. Similarly, $M_\Psi^\dagger$ takes in a state in $\mathcal{H}_{\mathcal{X}_R}$, time evolves by $\frac{\tilde{\beta}_R}{2}$, applies $O_\Psi^\dagger$

and finally time evolves by $\frac{\tilde{\beta}_L}{2}$ to define a state in $\mathcal{H}_{\mathcal{X}_L}$. In this notation the above “glued” boundary condition is produced by $M_\Psi M_\Psi^\dagger$.

To make progress let us insert the identity to write $M_\Psi M_\Psi^\dagger = M_\Psi \mathbb{1}_{\mathcal{H}_{\mathcal{X}_L}} M_\Psi^\dagger$. We resolve this insertion of the identity in a basis “conjugate” to the one used for the trace over $\mathcal{H}_{\mathcal{X}_L}$, defined as follows. Given a basis set $\{|a\rangle_L\}$ defined by operator insertions O_a on some cut manifold $\mathcal{M}_a$ with cut $\mathcal{X}$, the conjugate basis $\{|\tilde{a}\rangle_L\}$ is defined by placing the operator $O_a^\dagger$ in place of O_a , hence $|\tilde{a}\rangle \equiv |a^*\rangle$. Note that if $\{|a\rangle_L\}$ forms a basis then so must $\{|\tilde{a}\rangle_L\}$. Resolving the identity in this conjugate basis we obtain:

$$M_\Psi \mathbb{1}_{\mathcal{H}_{\mathcal{X}_L}} M_\Psi^\dagger = \lim_{m \rightarrow -1} (G^m)_{L,b^*a^*} M_\Psi |b^*\rangle_L \langle a^*|_L M_\Psi^\dagger. \quad (4.5)$$

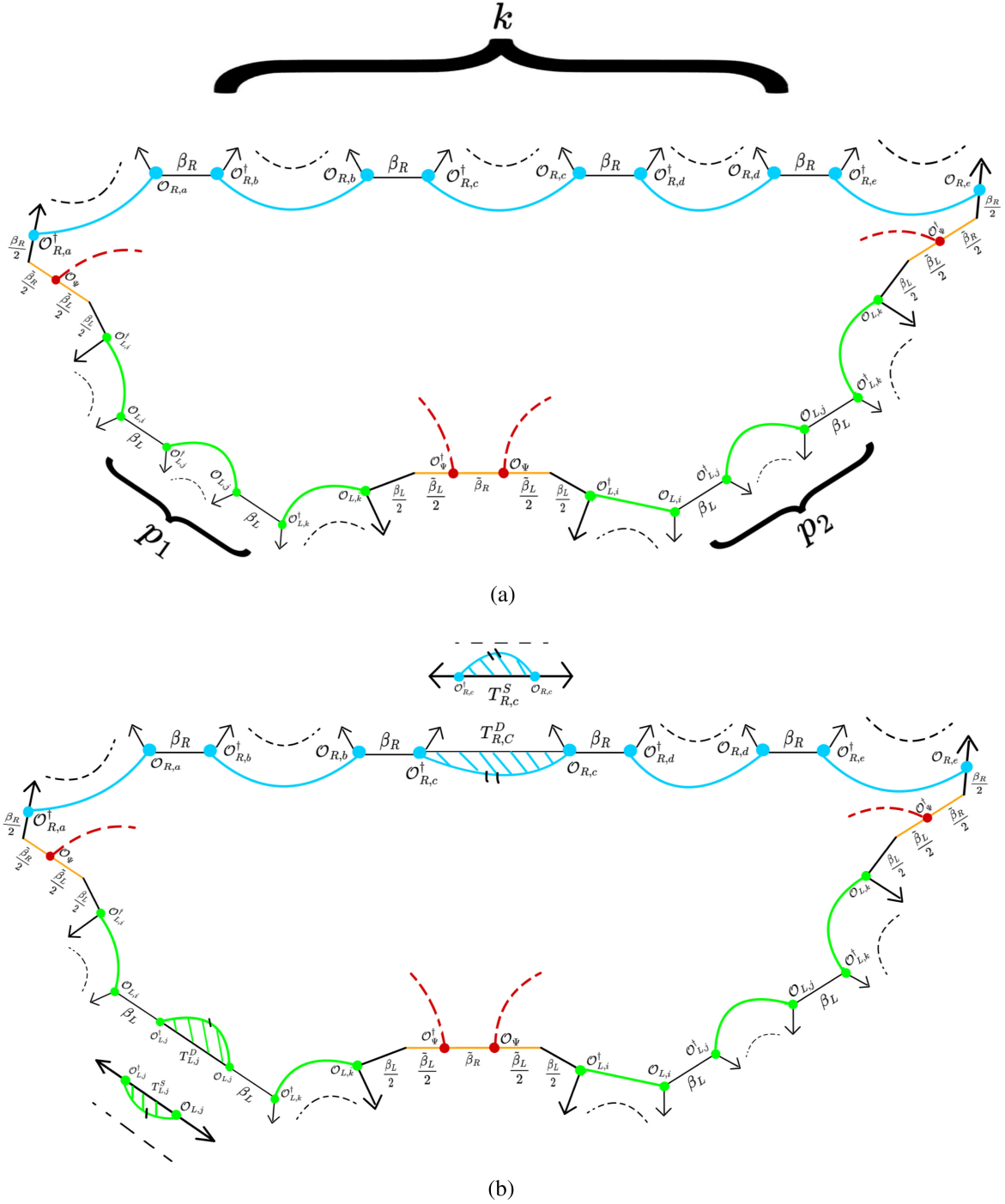


FIG. 11. (a) The maximally connected saddles to $\overline{(G^k)_{R,cd} \langle d | \prod_{i=1}^n (G^{p_i})_{L,a_i b_i} \langle b_i |_L \Psi \rangle \langle \Psi | a_i \rangle_L | c \rangle_R}$ connect all the boundaries in a single bulk. (b) These saddles are constructed by gluing a shell strip into the central disk along the shell-world volume. This is the case for all the shells, but is depicted here for one L and one R shell for clarity. The red dotted lines convey that this construction is agnostic about the dynamics of the O_Ψ insertions.

Now, for any two states $|m\rangle_R, |n\rangle_R \in \mathcal{H}_{\mathcal{X}_R}$ we have $\langle m |_R M_\Psi | b^* \rangle_L = \langle m |_R \langle b |_L | \Psi \rangle$ and similarly $\langle a^* |_L M_\Psi^\dagger | n \rangle_R = \langle \Psi | a \rangle_L \langle n |_R$. By definition we also have $G_{b^* a^*} = \langle b^* | a^* \rangle = \langle a | b \rangle = G_{ab}$ and therefore the matrix elements of (4.4)

and (4.5) are equal. It follows that $\rho_R = M_\Psi M_\Psi^\dagger$ and hence the reduced density matrix on $\mathcal{H}_{\mathcal{X}_R}$ is indeed obtained by sewing the ket and bra together, showing requirement (1). This allows us to use n -fold concatenation

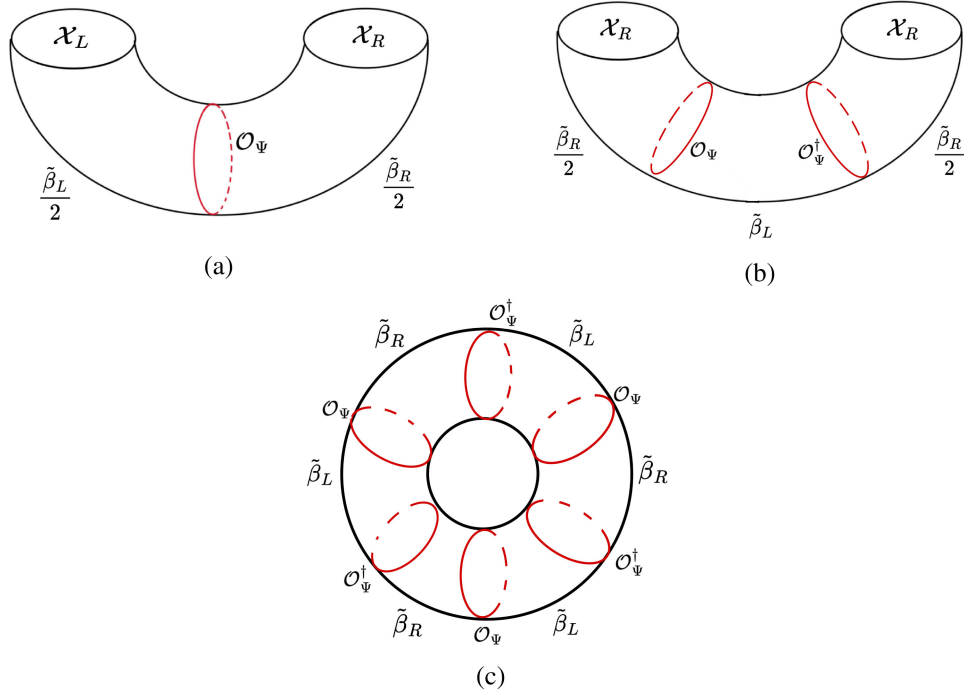


FIG. 12. Path integral prescription for calculating the n th Rényi entropy (4.3) of in QFT for a state living in $\mathcal{H}_{\text{QFT}} \otimes \mathcal{H}_{\text{QFT}}$: (a) Start with the condition defining a two-sided state $|\Psi\rangle$ which alternatively can be viewed as the operator $M_\Psi: \mathcal{H}_{\mathcal{X}_L} \rightarrow \mathcal{H}_{\mathcal{X}_R}$. (b) The reduced density matrix $\rho_R = \text{Tr}_{\mathcal{H}_{\mathcal{X}_L}}(|\Psi\rangle\langle\Psi|)$ is constructed by gluing the $|\Psi\rangle$ ket to the bra $\langle\Psi|$ along the cut $\mathcal{X}_L$. This can alternatively be viewed as the operator $M_\Psi M_\Psi^\dagger: \mathcal{H}_{\mathcal{X}_R} \rightarrow \mathcal{H}_{\mathcal{X}_R}$. (c) $\text{Tr}_{\mathcal{H}_{\mathcal{X}_R}}(\rho_R^n)$ is then given by gluing n copies of ρ_R to each other along the $\mathcal{X}_R$ cuts and the gluing the final two $\mathcal{X}_R$ cuts together to obtain a periodic path integral.

of the sewn boundary condition $\rho_R^n = (M_\Psi M_\Psi^\dagger)^n$ instead of the more involved boundary condition $\rho_R^n = \lim_{m_1, \dots, m_n \rightarrow -1} \times \prod_{i=1}^n (G^{m_i})_{L, a_i b_i} \langle b_i | \Psi \rangle \langle \Psi | a_i \rangle_L$ to compute Rényi entropies.

Factorization sews the trace. Having shown that the trace over $\mathcal{H}_{\mathcal{X}_L}$ sews the bra and ket together into

$M_\Psi M_\Psi^\dagger$ we can compute the trace as $\text{Tr}_{\mathcal{H}_{\mathcal{X}_R}}(\rho_R^n) = \text{Tr}_{\mathcal{H}_{\mathcal{X}_R}}(\prod_{i=1}^n M_{\Psi, i} M_{\Psi, i}^\dagger)$. We can rewrite this by inserting the identity on $\mathcal{H}_{\mathcal{X}_L}$ in between the k -th $M_\Psi M_\Psi^\dagger$ insertion, resulting in $\text{Tr}_{\mathcal{H}_{\mathcal{X}_R}}(\rho_R^n) = \text{Tr}_{\mathcal{H}_{\mathcal{X}_R}}(M_{\Psi, 1} M_{\Psi, 1}^\dagger \cdots M_{\Psi, k} \mathbb{1}_{\mathcal{H}_{\mathcal{X}_L}} M_{\Psi, k}^\dagger \cdots M_{\Psi, n} M_{\Psi, n}^\dagger)$. Resolving the trace in

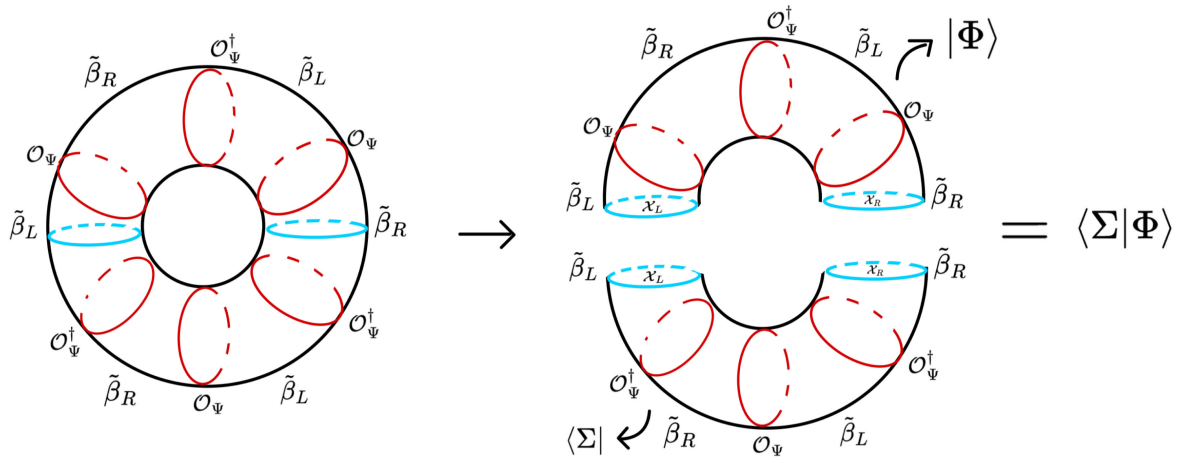


FIG. 13. The path integral $Z_0(\rho_\Psi^n)(\beta_L, \beta_R)$ can be rewritten as the overlap $\langle \Sigma | \Phi \rangle$ by cutting it into two parts along cuts $\mathcal{X}_L$ and $\mathcal{X}_R$.

the $\mathcal{H}_{\mathcal{X}_R}$ conjugate basis we obtain

$$\begin{aligned} \text{Tr}_{\mathcal{H}_{\mathcal{X}_R}}(\rho_R^n) &= \lim_{x,y \rightarrow -1} (G^x)_{R,a^*b^*} (G^y)_{L,st} \\ &\times \langle b^* |_R M_{\Psi,1} M_{\Psi,1}^\dagger \cdots M_{\Psi,k} | s \rangle_L \\ &\times \langle t |_L M_{\Psi,k}^\dagger \cdots M_{\Psi,n} M_{\Psi,n}^\dagger | a^* \rangle_R. \end{aligned} \quad (4.6)$$

Note that $M_{\Psi,1} M_{\Psi,1}^\dagger \cdots M_{\Psi,k}$ and $M_{\Psi,k}^\dagger \cdots M_{\Psi,n} M_{\Psi,n}^\dagger$ can be seen as a bra $\langle \Sigma |$ and ket $|\Phi \rangle$ on $\mathcal{H}_{\mathcal{X}_L \cup \mathcal{X}_R}$ respectively. In this notation the path integral with the boundary condition defining $Z_\circ(\rho_\Psi^n)(\tilde{\beta}_L, \tilde{\beta}_R)$ computes the overlap between these states (Fig. 13)

$$Z_\circ(\rho_\Psi^n)(\tilde{\beta}_L, \tilde{\beta}_R) = \langle \Sigma | \Phi \rangle. \quad (4.7)$$

Furthermore we have $\langle b^* |_R M_{\Psi,1} M_{\Psi,1}^\dagger \cdots M_{\Psi,k} | s \rangle_L = \langle \Sigma | s \rangle_L | b \rangle_R$ and $\langle t |_L M_{\Psi,k}^\dagger \cdots M_{\Psi,n} M_{\Psi,n}^\dagger | a^* \rangle_R = \langle t |_L \langle a |_R | \Phi \rangle$ and $G_{R,a^*b^*} = G_{R,ba}$. Using this we obtain:

$$\begin{aligned} \text{Tr}_{\mathcal{H}_{\mathcal{X}_R}}(\rho_R^n) &= \lim_{x,y \rightarrow -1} G_{R,ba} (G^y)_{L,st} \langle \Sigma | s \rangle_L | b \rangle_R \langle t |_L \langle a |_R | \Phi \rangle \\ &= \langle \Sigma | \mathbb{1}_{\mathcal{X}_L} \otimes \mathbb{1}_{\mathcal{X}_R} | \Phi \rangle = \langle \Sigma | \Phi \rangle, \end{aligned} \quad (4.8)$$

where the last equality follows from the factorization (2.1) of the two-boundary Hilbert space $\mathcal{H}_{\mathcal{X}_L \cup \mathcal{X}_R}$. As in Sec. III A, once factorization (2.1) is granted this last step is a consequence of linear algebra; the gravitational content resides entirely in the factorization result of [6], and this subsection should not be read as an independent gravitational derivation. Hence the two distinct boundary conditions (4.2) and $Z_\circ(\rho_\Psi^n)(\tilde{\beta}_L, \tilde{\beta}_R)$ for the path integral computing Rényi entropy define the same gravitational path integral and (4.3) holds in the fine-grained theory. Here we have shown requirement (2); i.e., for any operator $\mathcal{U}: \mathcal{H}_{\mathcal{X}_R} \rightarrow \mathcal{H}_{\mathcal{X}_R}$ the role of the trace $\text{Tr}_{\mathcal{H}_{\mathcal{X}_R}}(\mathcal{U})$ is to sew the boundary condition defining the operator together along the two $\mathcal{X}_R$ cuts. This sewing is exactly what we expect if there were a quantum theory living on the boundary that is dual to the bulk gravitational system. Interestingly, the derivation given here relies only on the factorization of the two boundary Hilbert space, which was shown in [6] to hold in asymptotically flat quantum gravity too.

B. Explicitly tracing out a universe

We now use the single-sided shell basis (see Sec. III B) to explicitly evaluate the traces in (4.2). In particular, we apply the toolkit of [5] to show that $\overline{\mathcal{A} - \mathcal{B}} = 0$ and $\overline{(\mathcal{A} - \mathcal{B})^2} = 0$ with $\mathcal{A} = \text{Tr}_{\mathcal{H}_{\mathcal{X}_R}}(\rho_R^n)$ and $\mathcal{B} = Z_\circ(\rho_\Psi^n)(\tilde{\beta}_L, \tilde{\beta}_R)$; together these give the fine-grained equality $\mathcal{A} = \mathcal{B}$, that is (4.3). The boundary condition for ρ_R^n consists of concatenating n copies of the ρ_R boundary condition as in Fig. 10(a). The

trace over $\mathcal{H}_R$ consists of adjoining an R shell state boundary to each of the two ends of ρ_R^n , thereby forming a “complete” boundary condition. These R shell insertions form an index loop with the k R -shell boundary insertions from the $(G^k)_R$ insertion, see Fig. 10(b).

The averaged equality. We consider the $\kappa_L, \kappa_R \rightarrow \infty$ limit in which the single-sided shell states span. As explained in [5] in these limits only geometries that do not break any index loops in (4.2) contribute. The mechanism is the same as for the thermal trace in Sec. III B; a geometry that breaks an index loop carries fewer free shell-flavor sums than one that preserves it, and is therefore suppressed by powers of $1/\kappa$ in the spanning limit. The surviving geometries are those in which each shell world-line is contracted with its conjugate, so that the L - and R -shell index loops of (4.2) are realized in the bulk by a single connected geometry. In the relevant saddle points for (4.2) at a given $\{k, p_i\}$, the shell index loops are reflected in the bulk by a loop of boundary and shell world volume segments [Fig. 11(a)]. These saddle points to (4.2) are constructed analogously to the wormhole saddles in Sec. III B; that is, we glue the interior portion of a disk with shell and $\mathcal{O}_\Psi, \mathcal{O}_\Psi^\dagger$ operator insertions on the asymptotic boundary to $k + \sum_{i=1}^n p_i$ shell strips. Each shell strip carries a single shell propagating between an $\mathcal{O}_S$ and an $\mathcal{O}_S^\dagger$ insertion; gluing it to the disk along that world volume and discarding the shell homology region on either side [Fig. 11(b)] contracts one shell index loop, so that the $k + \sum_i p_i$ strips realize all the index contractions of (4.2). The L shells propagate through for a boundary time $T_{L,i}$ on the disk boundary and a time $T_{L,i}^S$ on the strip boundary, and similarly for the R shells. The Israel junction conditions glue these world volumes and fix the propagation times $\Delta T_{L,i,j}, \Delta T_{R,j}$ and their strip counterparts, exactly as for the wormhole saddles of Sec. III B. Apart from the shell insertions, the operator insertions on the boundary of the disk are similar to that of $Z_\circ(\rho_\Psi^n)(\tilde{\beta}_L, \tilde{\beta}_R)$, where instead the $i \neq n$ th $\mathcal{O}_\Psi, \mathcal{O}_\Psi^\dagger$ operator pair is separated by a boundary time $\tilde{\beta}_L + (p_i + 1)\beta_L + \sum_{j=1}^{p_i+1} \Delta T_{L,i,j}$ and $\mathcal{O}_{\Psi,1}, \mathcal{O}_{\Psi,n}^\dagger$ separated by $\tilde{\beta}_R + (k + 1)\beta_R + \sum_{j=1}^{k+1} \Delta T_{R,j}$.

In the large-shell mass limit, all the shell propagation times $\Delta T_i, \Delta T_i^S \rightarrow 0$ go to zero, the shell homology regions on the disk and strips pinch off and the shells contribute universally to the action. The contribution from all the L shells in this limit is $\prod_{i=1}^n \prod_{j=1}^{p_i+1} Z_{m_j,L}$ and that for the R shells $\prod_{l=1}^{k+1} Z_{m_l,R}$. As $\Delta T_i^S \rightarrow 0$ the contribution from each strip is simply that of the strip of length zero $\tilde{S}(0)$, resulting in a factor of $\tilde{S}(0)^{k+1} \prod_{i=1}^n \tilde{S}(0)^{p_i+1}$. These shell and strip terms all cancel after normalizing the shell states to unit norm [see Eq. (3.9)]. Explicitly, each normalized shell carries a factor $1/\sqrt{Z_m \tilde{Z}(\beta) \tilde{S}(0)}$ from (3.9), and there are precisely as many such factors as there are shell and strip insertions, so the product of Z_m and $\tilde{S}(0)$ factors is

canceled. Lastly, the contribution to the action from the central disk section in this limit is that of the saddle geometries satisfying the periodic asymptotic boundary condition where the $i \neq n$ th $\mathcal{O}_\Psi, \mathcal{O}_\Psi^\dagger$ operator pair is separated by a boundary time $\tilde{\beta}_L + (p_i + 1)\beta_L$ and $\mathcal{O}_{\Psi,1}, \mathcal{O}_{\Psi,n}^\dagger$ separated by $\tilde{\beta}_R + (k + 1)\beta_R$. In the $k, p_1, \dots, p_n \rightarrow -1$ limit this disk boundary condition becomes precisely $Z_\circ(\rho_\Psi^n)(\tilde{\beta}_L, \tilde{\beta}_R)$ [see Fig. 11(b)], so that $\bar{\mathcal{A}} = \bar{\mathcal{B}}$.

The variance and cross-replica wormholes. To upgrade the equality of averages to the fine-grained equality (4.3) we must also show $\overline{(\mathcal{A} - \mathcal{B})^2} = 0$. Expanding the square,

$$\overline{(\mathcal{A} - \mathcal{B})^2} = \overline{\mathcal{A}^2} - 2\overline{\mathcal{A}\mathcal{B}} + \overline{\mathcal{B}^2}, \quad (4.9)$$

so it suffices to show that $\overline{\mathcal{A}^2}$, $\overline{\mathcal{A}\mathcal{B}}$ and $\overline{\mathcal{B}^2}$ all agree. Each is a sum of a disconnected piece—the product of the two means, which already agree by the result above—and connected pieces in which bulk geometries link the two replicas. It is therefore enough to match these

connected, wormhole contributions across the three quantities.

Consider $\overline{\mathcal{A} \times \mathcal{B}}$. Whether connected saddles linking the two replicas exist depends on the operator $\mathcal{O}_\Psi$ defining the state. If the quantity $\overline{\mathcal{B} \times \mathcal{B}} = \overline{Z_\circ \times Z_\circ}$ has no connected saddle, then neither does $\overline{\mathcal{A} \times \mathcal{B}}$, and the three terms in (4.9) are trivially equal. When connected saddles do exist, they are built exactly as the single-replica saddles above were: by gluing the shell strips of the $\mathcal{A}$ replica into the fully-connected saddles of $\overline{Z_\circ \times Z_\circ}$ along the shell world volumes (this is the analog of the construction in [5,6]; see for example Fig. 18 in [6]). The strips again shift the boundary times of the $\mathcal{A}$ replica's $Z_\circ$ factor by the shell preparation temperatures and propagation times. In particular, if there exist fully connected contributions to $\overline{Z_\circ(\rho_\Psi^n)(\tilde{\beta}_L, \tilde{\beta}_R) \times Z_\circ(\rho_\Psi^n)(\tilde{\beta}_L, \tilde{\beta}_R)}$ then the additional wormhole saddles relevant for $\overline{\text{Tr}_{\mathcal{H}_{\chi_R}}(\rho_R^n) \times Z_\circ(\rho_\Psi^n)(\tilde{\beta}_L, \tilde{\beta}_R)}$ are constructed by gluing shell strips into the fully-connected saddles to

$$Z_\circ(\rho_\Psi^n) \left(\tilde{\beta}_L + (p_i + 1)\beta_L + \sum_{j=1}^{p_i+1} \Delta T_{L,i,j}, \tilde{\beta}_R + (k + 1)\beta_R + \sum_{j=1}^{k+1} \Delta T_{R,j} \right) \times Z_\circ(\rho_\Psi^n)(\tilde{\beta}_L, \tilde{\beta}_R). \quad (4.10)$$

This time evolved $Z_\circ$ boundary condition accommodates the propagation times of the shells, which go to zero as $m_i \rightarrow \infty$, and the L, R shell inverse preparation temperatures β_L, β_R . Taking the large-shell mass limit exactly as for the mean sends the propagation times to zero and cancels the strip and shell factors against the unit-norm normalization, so the shifted arguments of the first $Z_\circ$ in (4.10) reduce to $\tilde{\beta}_L, \tilde{\beta}_R$. In the $k, p_1, \dots, p_n \rightarrow -1$ limit each of the fully connected saddles to (4.10) limits to the connected contribution to $\overline{Z_\circ(\rho_\Psi^n)(\tilde{\beta}_L, \tilde{\beta}_R) \times Z_\circ(\rho_\Psi^n)(\tilde{\beta}_L, \tilde{\beta}_R)} = \overline{\mathcal{B}^2}$. The identical construction with the second replica also prepared as a shell-trace produces the connected saddles of $\overline{\mathcal{A}^2}$, which limit to the same object. Hence $\overline{\mathcal{A}^2} = \overline{\mathcal{A}\mathcal{B}} = \overline{\mathcal{B}^2}$, and $\overline{(\mathcal{A} - \mathcal{B})^2} = 0$. This shows $\text{Tr}_{\mathcal{H}_{\chi_R}}(\rho_R^n) = Z_\circ(\rho_\Psi^n)(\tilde{\beta}_L, \tilde{\beta}_R)$ in the fine-grained theory, at least within the saddle-point approximation.

We may again extend this analysis beyond the leading saddle. As with the thermal trace, the cleanest statement is the factorization argument outlined above; we showed there that $\text{Tr}_{\mathcal{H}_{\chi_R}}(\rho_R^n)$ and $\langle \Sigma | \mathbb{1}_{\chi_L} \otimes \mathbb{1}_{\chi_R} | \Phi \rangle$ represent path integrals with the same boundary conditions [see the discussion around (4.7)], and that, once the two-boundary Hilbert space factorizes, $\text{Tr}_{\mathcal{H}_{\chi_R}}(\rho_R^n) = \langle \Sigma | \Phi \rangle = Z_\circ(\rho_\Psi^n)(\tilde{\beta}_L, \tilde{\beta}_R)$ follows purely algebraically; this is therefore valid for

the full path integral whenever it is suitably defined, independently of any saddle expansion. The explicit shell-basis evaluation above provides an independent check, which can be extended beyond the leading saddle by the same heuristic argument as in Sec. III B; the saddles of the two sides are in one-to-one correspondence, and in the large shell-mass limit the strips decouple while the heavy shells localize near the asymptotic boundary, where with standard boundary conditions the fluctuations they couple to are normalizable and subleading, so that to the order we work the perturbative corrections reside in the central region and match those of $Z_\circ(\rho_\Psi^n)$ order by order. As before, we do not make this second argument precise—it rests on the suppression of the near-boundary fluctuations in the large-mass limit—and we rely on the factorization argument for the clean full-path-integral statement.

V. SUMMARY AND DISCUSSION

In this paper we have explained why the Euclidean path integral with one asymptotic boundary and periodic time computes a thermal trace over the gravity Hilbert space, as Gibbons and Hawking proposed. We have shown equality between these *a priori* different gravitational path integrals in two ways. First we cut the Gibbons-Hawking path integral in half to write it as the overlap of a two-boundary gravity state and inserted a resolution of the two-boundary

identity in between the bra and ket. We showed that this expression is equal to the single-boundary thermal trace if the two-boundary Hilbert space factorizes into copies of the single-boundary theories, which was established in [6]. This argument is completely general and did not rely on a particular form of the gravitational action. Our second argument used a particular basis for the single-boundary Hilbert space to compute the thermal trace explicitly, and again found equality between the two prescriptions. Interestingly, in either method nonperturbative gravitational effects are crucial, suggesting the Gibbons-Hawking path integral does not equal the thermal trace in perturbative gravity.

While the expression $\text{Tr}_{\mathcal{H}_X}(e^{-\beta H}) = \langle \beta | \mathbb{1}_{\mathcal{X}_L} \otimes \mathbb{1}_{\mathcal{X}_R} | \beta \rangle$ derived around (3.4) shows that the factorization $\mathcal{H}_{\mathcal{X}_L \cup \mathcal{X}_R} = \mathcal{H}_{\mathcal{X}_L} \otimes \mathcal{H}_{\mathcal{X}_R}$ is sufficient for $\text{Tr}_{\mathcal{H}_X}(e^{-\beta H_X}) = Z(\beta)$ to be true, it does not establish that it is necessary. Indeed, in this section we have shown $\text{Tr}_{\mathcal{H}_X}(e^{-\beta H_X}) = Z(\beta)$ without appealing to factorization at all. Similarly, [5] showed the fine-grained equality,

$$\text{Tr}_{\mathcal{H}_{LR}}(e^{-H_L \beta_1} e^{-H_R \beta_2}) = Z(\beta_1) \times Z(\beta_2), \quad (5.1)$$

by tracing over a basis of two-sided shell states, also without requiring factorization. The derivation of (5.1) in [5] and the derivation of (3.12) above relied instead on the existence of wormhole geometries that separately connect two-boundary or one-boundary states. By contrast, factorization of the two-boundary Hilbert space (2.1) followed from the existence of a *different* nonperturbative effect—“mixed” wormholes connecting two-boundary states to the one-boundary states [6]. Thus, $\text{Tr}_{\mathcal{H}_X}(e^{-\beta H_X}) = Z(\beta)$ can follow without factorization of the two-boundary Hilbert space if there is a consistent theory of gravity in which wormholes connecting only one- or two- boundary states exist but “mixed” wormholes do not. It would be interesting therefore to understand whether or not existence of the first kind of wormhole requires the second for consistency of the theory.

Our analysis in fact suggests the opposite conclusion in theories where the two-boundary Hilbert space does *not* factorize into single-boundary factors, such as matter-free JT gravity and pure 3D gravity. The equality of the Gibbons-Hawking path integral with the single-boundary thermal trace rested on identifying $\mathbb{1}_{\mathcal{X}_L} \otimes \mathbb{1}_{\mathcal{X}_R}$ with $\mathbb{1}_{\mathcal{X}_L \cup \mathcal{X}_R}$, an identification that holds precisely when the two-boundary Hilbert space factorizes. When factorization fails these operators differ, and the Gibbons-Hawking path integral $\overline{\langle \beta | \beta \rangle} = \overline{\langle \beta | \mathbb{1}_{\mathcal{X}_L \cup \mathcal{X}_R} | \beta \rangle}$ no longer equals the single-boundary trace $\overline{\langle \beta | \mathbb{1}_{\mathcal{X}_L} \otimes \mathbb{1}_{\mathcal{X}_R} | \beta \rangle}$. In such theories, then, our results suggest that the Gibbons-Hawking path integral does not admit an interpretation as a thermal trace

over the single-boundary Hilbert space at all. This is consistent with the factorization puzzle of [38], in which the absence of a factorized Hilbert space obstructs a clean trace interpretation of the gravitational partition function. It would be interesting to understand what, if anything, the Gibbons-Hawking path integral does compute in these cases.

A corollary of our result is that the exponential of the Bekenstein-Hawking entropy does count the leading order microcanonical density of states of the gravity theory. We then used the fact that the single-sided shell states span, as shown in [6], to argue that the basis for the microcanonical Hilbert space consists solely of black hole states with the same exterior as the Gibbons-Hawking black hole saddle but with a nontrivial interior. Therefore the Euclidean path integral with one boundary and periodic time indeed counts the number of black hole microstates after microcanonical projection, a fact that was expected but not previously demonstrated through a direct calculation in gravity.

By similar methods we also arrived at a purely gravitational understanding of why the holographic replica trick is in fact evaluating the trace required for computing the Rényi entropy of gravity states entangled between two universes. As the two-boundary Hilbert space factorizes into single-boundary theories, we obtained the entanglement entropy by tracing out these factors individually. We showed equality with the holographic replica trick through a general factorization argument and also by explicitly tracing over the shell basis. We focused our analysis on entanglement between universes with disconnected boundaries. It would be interesting to see if these methods can be extended to tracing out a subregion of a connected boundary component, and to see how our construction is related to the algebraic understanding of gravitational entropy recently discussed in [39].

In this paper we made no assumptions regarding holographic duality and our arguments apply equally in asymptotically AdS and flat spacetime.

ACKNOWLEDGMENTS

T. Y. thanks the Peter Davies Scholarship for continued support. V. B. was supported in part by the DOE through Grant No. DE-SC0013528 and QuantISED Grant No. DE-SC0020360, and in part by the Eastman Professorship at Balliol College, University of Oxford. V. B. thanks Simon Ross and Sameer Murthy for discussions about the Gibbons-Hawking approach to black hole entropy.

DATA AVAILABILITY

The data that support the findings of this article are openly available [1].

- [1] J. D. Bekenstein, Black holes and entropy, *Phys. Rev. D* **7**, 2333 (1973).
- [2] J. M. Bardeen, B. Carter, and S. W. Hawking, The four laws of black hole mechanics, *Commun. Math. Phys.* **31**, 161 (1973).
- [3] S. W. Hawking, Particle creation by black holes, *Commun. Math. Phys.* **43**, 199 (1975); *Commun. Math. Phys.* **46**, 206 (1976).
- [4] G. W. Gibbons and S. W. Hawking, Action integrals and partition functions in quantum gravity, *Phys. Rev. D* **15**, 2752 (1977).
- [5] V. Balasubramanian and T. Yildirim, A nonperturbative toolkit for quantum gravity, [arXiv:2504.16986](https://arxiv.org/abs/2504.16986).
- [6] V. Balasubramanian and T. Yildirim, The nonperturbative Hilbert space of quantum gravity with one boundary, *J. High Energy Phys.* **03** (2026) 040.
- [7] M. Sasieta, Wormholes from heavy operator statistics in AdS/CFT, *J. High Energy Phys.* **03** (2023) 158.
- [8] V. Balasubramanian, A. Lawrence, J. M. Magan, and M. Sasieta, Microscopic origin of the entropy of black holes in general relativity, *Phys. Rev. X* **14**, 011024 (2024).
- [9] V. Balasubramanian, A. Lawrence, J. M. Magan, and M. Sasieta, Microscopic origin of the entropy of astrophysical black holes, *Phys. Rev. Lett.* **132**, 141501 (2024).
- [10] A. Climent, R. Emparan, J. M. Magan, M. Sasieta, and A. Vilar López, Universal construction of black hole microstates, *Phys. Rev. D* **109**, 086024 (2024).
- [11] V. Balasubramanian, B. Craps, J. Hernandez, M. Khramtsov, and M. Knysh, Counting microstates of out-of-equilibrium black hole fluctuations, *J. High Energy Phys.* **06** (2025) 083.
- [12] V. Keranen, H. Nishimura, S. Stricker, O. Taanila, and A. Vuorinen, Gravitational collapse of thin shells: Time evolution of the holographic entanglement entropy, *J. High Energy Phys.* **06** (2015) 126.
- [13] J. Chandra and T. Hartman, Coarse graining pure states in AdS/CFT, *J. High Energy Phys.* **10** (2023) 030.
- [14] J. Boruch, L. V. Iliesiu, G. Lin, and C. Yan, How the Hilbert space of two-sided black holes factorises, *J. High Energy Phys.* **06** (2025) 092.
- [15] V. Balasubramanian, B. Craps, J. Hernandez, M. Khramtsov, and M. Knysh, Factorization of the Hilbert space of eternal black holes in general relativity, *J. High Energy Phys.* **01** (2025) 046.
- [16] P. Saad, S. H. Shenker, and D. Stanford, JT gravity as a matrix integral, [arXiv:1903.11115](https://arxiv.org/abs/1903.11115).
- [17] P. Saad, Late time correlation functions, baby universes, and ETH in JT gravity, [arXiv:1910.10311](https://arxiv.org/abs/1910.10311).
- [18] D. Marolf and H. Maxfield, Transcending the ensemble: Baby universes, spacetime wormholes, and the order and disorder of black hole information, *J. High Energy Phys.* **08** (2020) 044.
- [19] J. de Boer, D. Liska, B. Post, and M. Sasieta, A principle of maximum ignorance for semiclassical gravity, *J. High Energy Phys.* **2024** (2024) 003.
- [20] J. de Boer, D. Liska, and B. Post, Multiboundary wormholes and OPE statistics, *J. High Energy Phys.* **10** (2024) 207.
- [21] G. Penington, S. H. Shenker, D. Stanford, and Z. Yang, Replica wormholes and the black hole interior, *J. High Energy Phys.* **03** (2022) 205.
- [22] 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.
- [23] D. Marolf, On the nature of ensembles from gravitational path integrals, *J. Phys. A* **58**, 035204 (2025).
- [24] S. Antonini, M. Sasieta, and B. Swingle, Cosmology from random entanglement, *J. High Energy Phys.* **11** (2023) 188.
- [25] B. Banihashemi and T. Jacobson, The enigmatic gravitational partition function, *Gen. Relativ. Gravit.* **57**, 43 (2025).
- [26] W. Israel, Singular hypersurfaces and thin shells in general relativity, *Nuovo Cimento B* **44S10**, 1 (1966); *Nuovo Cimento B* **48**, 463 (1967).
- [27] S. W. Hawking and D. N. Page, Thermodynamics of black holes in anti-de Sitter space, *Commun. Math. Phys.* **87**, 577 (1983).
- [28] H. Geng and Y. Jiang, Microscopic origin of the entropy of single-sided black holes, *J. High Energy Phys.* **04** (2025) 133.
- [29] J. M. Maldacena, Eternal black holes in anti-de Sitter, *J. High Energy Phys.* **04** (2003) 021.
- [30] A. Goel, H. T. Lam, G. J. Turiaci, and H. Verlinde, expanding the black hole interior: Partially entangled thermal states in SYK, *J. High Energy Phys.* **02** (2019) 156.
- [31] V. Balasubramanian, A. Kar, and T. Ugajin, Entanglement between two gravitating universes, *Classical Quantum Gravity* **39**, 174001 (2022).
- [32] V. Balasubramanian, A. Kar, and T. Ugajin, Entanglement between two disjoint universes, *J. High Energy Phys.* **02** (2021) 136.
- [33] V. Balasubramanian, A. Kar, and T. Ugajin, Islands in de Sitter space, *J. High Energy Phys.* **02** (2021) 072.
- [34] A. Lewkowycz and J. Maldacena, Generalized gravitational entropy, *J. High Energy Phys.* **08** (2013) 090.
- [35] S. Ryu and T. Takayanagi, Holographic derivation of entanglement entropy from AdS/CFT, *Phys. Rev. Lett.* **96**, 181602 (2006).
- [36] S. Ryu and T. Takayanagi, Aspects of holographic entanglement entropy, *J. High Energy Phys.* **08** (2006) 045.
- [37] V. E. Hubeny, M. Rangamani, and T. Takayanagi, A covariant holographic entanglement entropy proposal, *J. High Energy Phys.* **07** (2007) 062.
- [38] D. Harlow and D. Jafferis, The factorization problem in Jackiw-Teitelboim gravity, *J. High Energy Phys.* **02** (2020) 177.
- [39] E. Colafranceschi, X. Dong, D. Marolf, and Z. Wang, Algebras and Hilbert spaces from gravitational path integrals. Understanding Ryu-Takayanagi/HRT as entropy without AdS/CFT, *J. High Energy Phys.* **10** (2024) 063.