

Null states and time evolution in a toy model of black hole dynamics

Xi Dong ^a, Maciej Kolanowski ^{a,b}, Xiaoyi Liu ^a, Donald Marolf ^a
and Zhencheng Wang ^{a,c}

^a*Department of Physics, University of California,
Santa Barbara, CA 93106, U.S.A.*

^b*Institute of Theoretical Physics, Faculty of Physics, University of Warsaw,
Pasteura 5, 02-093 Warsaw, Poland*

^c*Department of Physics, University of Illinois Urbana-Champaign,
Urbana, IL 61801, U.S.A.*

E-mail: xidong@ucsb.edu, mkolanowski@ucsb.edu, xiaoyiliu@ucsb.edu,
marolf@ucsb.edu, zcwang1@illinois.edu

ABSTRACT: Spacetime wormholes can provide non-perturbative contributions to the gravitational path integral that make the actual number of states e^S in a gravitational system much smaller than the number of states e^{S_p} predicted by perturbative semiclassical effective field theory. The effects on the physics of the system are naturally profound in contexts in which the perturbative description actively involves $N = O(e^S)$ of the possible e^{S_p} perturbative states; e.g., in late stages of black hole evaporation. Such contexts are typically associated with the existence of non-trivial quantum extremal surfaces. However, by forcing a simple topological gravity model to evolve in time, we find that such effects can also have large impact for $N \ll e^S$ (in which case no quantum extremal surfaces can arise). In particular, even for small N , the insertion of generic operators into the path integral can cause the non-perturbative time evolution to differ dramatically from perturbative expectations. On the other hand, this discrepancy is small for the special case where the inserted operators are non-trivial only in a subspace of dimension $D \ll e^S$. We thus study this latter case in detail. We also discuss potential implications for more realistic gravitational systems.

KEYWORDS: AdS-CFT Correspondence, Black Holes, Models of Quantum Gravity

ARXIV EPRINT: [2405.04571](https://arxiv.org/abs/2405.04571)

Contents

1	Introduction	1
2	The Euclidean gravitational path integral and our topological model	3
2.1	The topological model	6
3	One-boundary Hilbert spaces and time evolution	8
3.1	The non-perturbative and perturbative Hilbert spaces	8
3.2	Formalism and notation	9
3.3	Operator insertions and time evolution	10
4	Large discrepancies for generic insertions	13
5	Long-time agreement for small-rank insertions	16
5.1	Eigenvalues and eigenvectors	17
5.2	Evolutions of a generic state	20
5.3	Degeneracies and approximate degeneracies	24
5.4	Finely-tuned states with large discrepancies	24
6	Discussion	25
A	Wishart distribution	29
A.1	Basic definitions	30
A.2	A few moments	30
A.3	Random projections	31

1 Introduction

Perturbative descriptions of quantum gravity are well-known to allow an arbitrarily large number of states inside a black hole with given area A . For example, a large collection of states is readily formed by starting with a black hole of very large area and then letting the black hole decay until it has area A . The tension between this fact and the finite Bekenstein-Hawking entropy forms the core of the so-called black hole information problem.

Recent work on gravitational path integrals and quantum extremal surfaces shows in a rather concrete way how this divergent perturbative result is transformed into a density of states that is finite at the non-perturbative level [1–4], and which is in particular given by the generalized entropy S_{gen} of an appropriate quantum extremal surface [5]. Important components are the inclusion of spacetime wormholes and the choice of a baby-universe superselection sector [6].

At least in simple contexts and with the use of these ingredients, the resulting non-perturbative path integral can be said to define a new inner product on the space of perturbative states. In particular, certain would-be perturbative states turn out to have

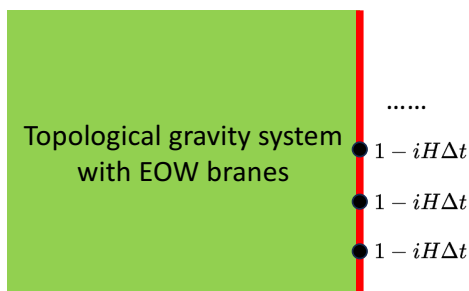


Figure 1. While the original dynamics of the topological model is trivial, we can force the system to evolve in time by inserting a large number of operators $(1 - iH\Delta t)$ with small Δt .

vanishing norm with respect to the full non-perturbative inner product [3, 6]. After taking the quotient of the perturbative Hilbert space by such null states, one finds the desired finite density of states; see also [7, 8].

This scenario provides an elegant solution to the state-counting issue. However, it is clearly of interest to understand any further implications for gravitational physics as well. Below, we use the 2d topological model of [6] (with end-of-the-world [EOW] branes) to study path integrals that compute time evolution in the presence of repeated boundary sources; see figure 1 above. It will be convenient to use the language of Euclidean path integrals, so that a real-time evolution is then generated by the insertion of complex sources. An immediate benefit of this approach will be to show that, in contrast with the concerns expressed in e.g. section 7.3 of [9], there is no tension between the presence of null states and unitary evolution. However, for general such sources, we will nevertheless find the non-perturbative results to differ markedly from perturbative expectations, even in contexts where we study only a small number of initial and final states.

This initial result then motivates us to further investigate a special class of boundary sources for which we expect non-perturbative contributions to be small. Namely, we consider operators of rank D on the perturbative Hilbert space with $D \ll d$, where d is the dimension of the non-perturbative Hilbert space. In such cases we find close agreement between the perturbative and non-perturbative path integrals, even when these integrals describe evolution over long times t of order d^α with $\alpha < \frac{1}{2}$. Furthermore, in this case, we also find there to be a map η from the perturbative Hilbert space $\mathcal{H}_p$ to the non-perturbative Hilbert space $\mathcal{H}_{np}$ that is approximately equivariant with respect to time evolution. In other words, if our path integral defines the perturbative time evolution operator e^{-iHt} and the non-perturbative time evolution operator $e^{-i\tilde{H}t}$, then acting on generic states $|\psi\rangle$ in $\mathcal{H}_p$ to good approximation we have

$$e^{-i\tilde{H}t}\eta|\psi\rangle \approx \eta e^{-iHt}|\psi\rangle. \quad (1.1)$$

Here and below we will use the term ‘generic states’ to mean states chosen without knowledge of the baby-universe superselection sector. Statements about generic states will hold with high probability in the limit where the dimension d of the non-perturbative Hilbert space is large.

We begin with a brief review in section 2 of the topological gravity model introduced in [6]. Section 3 then discusses the perturbative and non-perturbative Hilbert spaces $\mathcal{H}_p$

and $\mathcal{H}_{\text{np}}$ as well as the natural map η from $\mathcal{H}_{\text{p}}$ to $\mathcal{H}_{\text{np}}$. The large discrepancies between perturbative and non-perturbative time evolutions for general sources are studied in section 4, while the long-time agreement for repeated sources of small rank is derived in section 5. We close with further discussion in section 6 emphasizing possible implications for more realistic models of gravity.

2 The Euclidean gravitational path integral and our topological model

The Euclidean gravitational path integral computes correlation functions of observables by integrating over a set of fields Φ that satisfy given boundary conditions. Such fields are generally taken to include a spacetime metric. For asymptotically Anti-de Sitter (AdS) gravitational theories, the gravitational path integral is usually interpreted as computing partition functions $Z[J]$ for each boundary condition J defined by an asymptotically locally AdS (AlAdS) boundary. In the context of AdS/CFT, the quantity $Z[J]$ is dual to a corresponding CFT partition function.

We will denote bulk fields Φ satisfying the boundary condition J by writing $\Phi \sim J$. For boundary conditions on a disconnected boundary manifold with n connected components, and with boundary condition J_i on the i th component, we use the following notation for the Euclidean gravitational path integral:

$$\langle Z[J_1] \cdots Z[J_n] \rangle := \int_{\Phi \sim J} \mathcal{D}\Phi e^{-S[\Phi]}. \quad (2.1)$$

Note that the description on the right-hand-side depends only on the union of the asymptotic boundaries and is thus manifestly independent of the numbering assigned to the various connected components of the boundary. As a result, such correlation functions are invariant under permuting the boundary conditions J_i ; e.g.,

$$\langle Z[J_1] Z[J_2] Z[J_3] \rangle = \langle Z[J_1] Z[J_3] Z[J_2] \rangle = \langle Z[J_3] Z[J_2] Z[J_1] \rangle, \text{ etc.} \quad (2.2)$$

A feature of the gravitational path integral is that it generally does not factorize over disconnected boundaries:

$$\langle Z[J_1] Z[J_2] \rangle \neq \langle Z[J_1] \rangle \langle Z[J_2] \rangle. \quad (2.3)$$

The difference between the two sides arises from the contributions of bulk spacetime wormholes that connect otherwise-separate asymptotic boundaries. This feature is inconsistent with expectations from AdS/CFT since CFT partition functions factorize on any disconnected manifold. However, it is expected that gravitational path integrals that fail to factorize can be instead interpreted as describing averages over a non-trivial ensemble of dual boundary theories; see e.g. [10, 11].

When the above ensemble structure arises, it can be seen directly in the bulk description. There it is related to the existence of baby universes [12, 13]. Such baby universes are necessarily a part of any bulk theory in which the Euclidean gravitational path integral sums over topologies. The baby universe states then span a subspace $\mathcal{H}_{\text{BU}}$ of the full quantum gravity Hilbert space, where states in $\mathcal{H}_{\text{BU}}$ can be obtained by slicing open gravitational

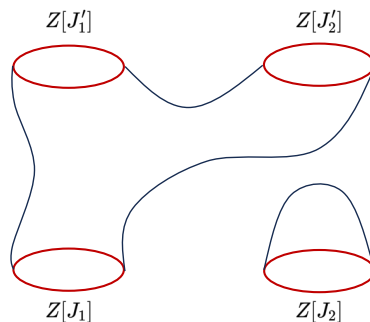


Figure 2. A particular contribution to the path integral that computes $\langle Z[J'_1]Z[J'_2]|Z[J_1]Z[J_2]\rangle$.

path integrals on slices that do not intersect any asymptotic boundaries. Our description of the above connections will follow [6], which emphasizes a basis of baby universe states that may be labeled

$$|Z[J_1]Z[J_2]\cdots Z[J_n]\rangle, \quad (2.4)$$

where $J_1, \dots, J_n$ again represent a set of closed asymptotic boundaries on which given boundary conditions are satisfied. There is in particular a ‘no-boundary Hartle-Hawking state’ $|\text{HH}\rangle$ associated with choosing $n = 0$. The inner product between two such states is then defined to be

$$\langle Z[J'_1]Z[J'_2]\cdots Z[J'_m]|Z[J_1]Z[J_2]\cdots Z[J_n]\rangle = \langle Z[J'_1]^*Z[J'_2]^*\cdots Z[J'_m]^*|Z[J_1]Z[J_2]\cdots Z[J_n]\rangle, \quad (2.5)$$

where $*$ denotes an appropriate anti-linear CPT-conjugation operation on the space of boundary conditions; see e.g. figure 2. One thinks of these boundaries as lying in the asymptotic Euclidean past or future so that space has no boundary at any finite time. We will assume the above inner product to satisfy reflection positivity:

$$\|\Psi\|^2 := \langle \Psi|\Psi\rangle \geq 0 \text{ for all } |\Psi\rangle = \sum_{i=1}^N c_i |Z[J_{i,1}]\cdots Z[J_{i,m_i}]\rangle. \quad (2.6)$$

The invariance of correlation functions with respect to permutations of the J_i (described under (2.1)) then implies the states (2.4) to be similarly invariant under such permutations of their $Z[J_i]$ labels; e.g. $|Z[J_1]Z[J_2]\rangle = |Z[J_2]Z[J_1]\rangle$.

A state of the form (2.4) can also be thought of as the result of acting with operators $\widehat{Z[J_i]}$ on the no-boundary Hartle-Hawking state $|\text{HH}\rangle$. In fact, we may define the $\widehat{Z[J_i]}$ by the relations:

$$|Z[J_1]Z[J_2]\cdots Z[J_n]\rangle = \widehat{Z[J_1]}\widehat{Z[J_2]}\cdots\widehat{Z[J_n]}|\text{HH}\rangle. \quad (2.7)$$

Since changing the order of operators does not alter the state, all of the $\widehat{Z[J]}$ operators commute with each other when acting on the dense domain spanned by states of the form (2.4). Furthermore, since (2.5) implies that the adjoints satisfy $\widehat{Z[J]}^\dagger = \widehat{Z[J^*]}$ on the given states, these operators also commute with all of the $\widehat{Z[J]}^\dagger$ operators when acting on such states.

As a result, if we assume that such results extend to a larger common domain on which all $\widehat{Z[J]} \pm \widehat{Z[J]}^\dagger$ satisfy the mathematical requirements to be fully self-adjoint (or anti self-adjoint) [14],¹ then the $\widehat{Z[J]}$ can all be diagonalized simultaneously. The simultaneous eigenvectors of all such operators are called α -states:

$$\widehat{Z[J]}|\alpha\rangle = Z_\alpha[J]|\alpha\rangle. \quad (2.8)$$

By inserting complete sets of α states, we find an ensemble structure:

$$\begin{aligned} \langle Z[J_1] \cdots Z[J_n] \rangle &= \sum_{\alpha_0, \alpha_1, \dots, \alpha_n} \langle \text{HH} | \alpha_0 \rangle \langle \alpha_0 | \widehat{Z[J_1]} | \alpha_1 \rangle \cdots \langle \alpha_{n-1} | \widehat{Z[J_n]} | \alpha_n \rangle \langle \alpha_n | \text{HH} \rangle \\ &= \mathfrak{Z} \sum_{\alpha} p_{\alpha} Z_{\alpha}[J_1] \cdots Z_{\alpha}[J_n], \end{aligned} \quad (2.9)$$

where $\mathfrak{Z} = \langle \text{HH} | \text{HH} \rangle$ is the norm of the Hartle-Hawking state, and $p_{\alpha} = \frac{|\langle \text{HH} | \alpha \rangle|^2}{\langle \text{HH} | \text{HH} \rangle}$ is the probability to measure the state $|\text{HH}\rangle$ to be in the state $|\alpha\rangle$. We also see that correlation functions factorize in α states. For example, when the α -states are normalizable we have

$$\langle \alpha | \widehat{Z[J_1]} \widehat{Z[J_2]} | \alpha \rangle = \langle \alpha | \widehat{Z[J_1]} | \alpha \rangle \langle \alpha | \widehat{Z[J_2]} | \alpha \rangle = Z_{\alpha}[J_1] Z_{\alpha}[J_2]. \quad (2.10)$$

We emphasize that a more standard AdS/CFT scenario, in which the bulk path integral respects factorization, can also be described in the above language. In that case, one simply finds that there is only one α -sector, so that the ensemble becomes trivial. The interesting consequence is then that there can be only one state in the baby universe Hilbert space $\mathcal{H}_{\text{BU}}$. Assuming that it is not a null state, the Hartle-Hawking no-boundary state $|\text{HH}\rangle$ is then non-perturbatively equivalent to the unique α -state.

Other sectors of the quantum gravity Hilbert space can be generated by cutting open the gravitational path integral along slices that intersect asymptotic boundaries on some non-trivial codimension-2 surface Σ . The full quantum gravity Hilbert space can be decomposed into sectors labelled by geometry (and perhaps other sources) on such Σ so that we may write

$$\mathcal{H}_{\text{QG}} = \bigoplus_{\Sigma} \mathcal{H}_{\Sigma}, \quad (2.11)$$

where $\Sigma = \emptyset$ corresponds to $\mathcal{H}_{\text{BU}}$. A general Hilbert space $\mathcal{H}_{\Sigma}$ also admits a decomposition into different α sectors,

$$\mathcal{H}_{\Sigma} = \bigoplus_{\alpha} \mathcal{H}_{\Sigma}^{\alpha}. \quad (2.12)$$

In particular, the set defined by the labels α is the same for all Σ ; see [6] for details.

¹The reader should be aware that the mathematics literature contains examples where such extensions fail spectacularly; see e.g. [14] for an example where symmetric operators that commute on a common invariant dense domain are essentially self-adjoint — so that they have a unique self-adjoint extension to the full Hilbert space — but where their self-adjoint extensions nevertheless fail to commute. However, it is far from clear that such failures occur in interesting models of gravitational physics.

2.1 The topological model

We now review the two-dimensional topological gravity model introduced in [6] with k flavors of end-of-the-world-branes (EOW branes). This gives a concrete and solvable model that demonstrates the features mentioned above. The action for the model is given by

$$S(\mathcal{M}) = -S_0 \chi(\mathcal{M}) - S_\partial n(\mathcal{M}), \quad (2.13)$$

where the action is evaluated on a compact two-dimensional surface $\mathcal{M}$ with Euler characteristic $\chi(\mathcal{M}) = 2 - 2g(\mathcal{M}) - n(\mathcal{M})$ where $g(\mathcal{M})$ is the genus of $\mathcal{M}$ and $n(\mathcal{M})$ is the number of circular boundaries (including both those that are determined by the boundary conditions J_i and those that arise dynamically from summing over loops of end-of-the-world branes). The theory allows two independent parameters, S_∂ and S_0 , but only certain choices satisfy reflection positivity.² For simplicity, we will use the choice $S_\partial = S_0$ below.

Because we consider a model with EOW branes, the surfaces $\mathcal{M}$ over which we will sum are allowed to have additional so-called dynamical boundaries, which are boundaries described entirely by EOW branes *not* specified by the asymptotic boundary conditions. Such dynamical boundaries must be labeled by some flavor I of the EOW branes. Circular (S^1) dynamical boundaries of the same flavor are considered indistinguishable.

The path integral for the model sums over all diffeomorphism classes of two-dimensional oriented surfaces that satisfy given boundary conditions:

$$\int_{\Phi \sim J} \mathcal{D}\Phi e^{-S[\Phi]} := \sum_{\text{Surfaces } \mathcal{M} \sim J} \mu(\mathcal{M}) e^{-S[\mathcal{M}]}, \quad (2.14)$$

where $\mu(\mathcal{M}) = \frac{1}{\prod_g m_g! \prod_{I=1}^k n_I!}$ if $\mathcal{M}$ has m_g connected components of genus g that have only dynamical boundaries, and if $\mathcal{M}$ has n_I dynamical boundaries associated with EOW-brane flavor I . The factor n_I accounts for residual gauge symmetries under the diffeomorphism-invariance described in [6]. Note that all boundaries specified by asymptotic boundary conditions are treated as distinguishable. In other words, we do not quotient by diffeomorphisms that relate disconnected components of such boundaries.

Another result of including EOW branes is that there are two kinds of connected asymptotic boundary conditions, which are associated with the following two kinds of operators on $\mathcal{H}_{\text{BU}}$:

- $\hat{Z}$, which creates a (unoriented) circular non-dynamical boundary. In the notation introduced earlier we might have written $J = \bigcirc$ and called this operator $\widehat{Z[\bigcirc]}$.
- $(\widehat{\psi_J, \psi_I})$, which creates an oriented interval of non-dynamical boundary running from a label I to a label J . One may think of such boundaries as describing the creation of an EOW brane with flavor I at the initial endpoint, and then describing the annihilation of a brane with flavor J at the final endpoint. Surfaces $\mathcal{M}$ in the sum (2.14) are allowed only if each initial label I is connected to a final label of the same flavor I by a

²The possible choices are $e^{S_\partial - S_0} \in \mathbb{N}$ due to the observation of [6] that the line segment defines a projection with trace $e^{S_\partial - S_0}$ in one of the α -sectors and the observation of [15] that positive-definiteness of the inner product requires traces of projections to take values in $\mathbb{N} \cup \{+\infty\}$.

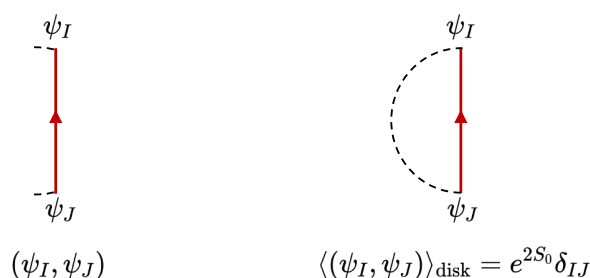


Figure 3. The labelled boundary interval (ψ_I, ψ_J) (left) and the disk contribution to $\langle(\psi_I, \psi_J)\rangle$ (right).

dynamical boundary segment. In the notation introduced earlier we might have written $J = I \rightarrow K$ and called this operator $Z[\widehat{I \rightarrow K}] := (\widehat{\psi_K, \psi_I})$; see figure 3.

Correlation functions of these operators can be computed from the gravitational path integral; see [6] for details. As described earlier, all of these operators commute, and they also commute with their adjoints. We may thus simultaneously diagonalize them. Their simultaneous eigenstates are the α -states.

After summing over all contributions, one finds that the Hartle-Hawking no-boundary state $|\text{HH}\rangle$ gives a Poisson distribution for the eigenvalues of $\hat{Z}$:

$$\mathfrak{Z}^{-1} \langle \hat{Z}^n \rangle = \sum_{d=0}^{\infty} d^n p_d(\lambda), \quad p_d(\lambda) = e^{-\lambda} \frac{\lambda^d}{d!}, \quad (2.15)$$

where $d \in \mathbb{N}$, and $\mathfrak{Z} = \langle 1 \rangle = e^\lambda$ for $\lambda = \frac{e^{2S_0}}{1 - e^{-2S_0}}$ is the contribution from summing over all vacuum bubbles. For $d \leq k$, we will find below that this Poisson distribution can also be interpreted as an ensemble over d -dimensional Hilbert spaces for the one-boundary sector with probabilities p_d .

Similar computations show that, within the sector where $Z = d$, the probabilities of the various eigenvalues $(\psi_J, \psi_I)_{d, \tilde{\alpha}}$ of $(\widehat{\psi_J, \psi_I})$ are given by a Wishart distribution. This means that we have³

$$(\psi_J, \psi_I)_{d, \tilde{\alpha}} = \frac{1}{d} \sum_{a=1}^d \bar{\psi}_J^a \psi_I^a, \quad (2.16)$$

where the ψ_I^a are independent random variables selected from a complex normal (Gaussian) ensemble with unit variance, and where the symbol $\tilde{\alpha}$ denotes the collection of ψ_I^a for all I, a . Some properties of Wishart distributions are given in appendix A.

A specific α -state is defined by choosing a particular instance $(d, \tilde{\alpha})$ of both the Poisson and Wishart distributions. Below, we will simply choose some fixed value of d by hand and then randomly draw the remaining variables $\tilde{\alpha}$ from the corresponding Wishart distribution.

³Note that our normalization differs from that of [6], by an overall factor $1/d$.

3 One-boundary Hilbert spaces and time evolution

The discussion in section 2 focussed on the baby universe Hilbert space $\mathcal{H}_{\text{BU}}$ in order to discuss α -sectors in their simplest context. However, our main interest in the current work will concern the space of one-boundary states; i.e., the space $\mathcal{H}_{\Sigma} = \bigoplus_{\alpha} \mathcal{H}_{\Sigma}^{\alpha}$ for the case where the codimension-2 boundary Σ is a single point. We will use the notation $\mathcal{H}_{\Sigma_1}^{\alpha}$ for the superselection sector of the one-boundary Hilbert space labelled by α . This Hilbert space and its perturbative analogue will be described in section 3.1. Section 3.2 clarifies the relation between the perturbative and non-perturbative Hilbert spaces and introduces notation to keep this manifest. Formalism for inserting operators into path integrals of the topological model (and thus inducing time evolution) will be described in section 3.3. This will set the stage for the analysis of such time evolutions in sections 4 and 5.

3.1 The non-perturbative and perturbative Hilbert spaces

An α -sector $\mathcal{H}_{\Sigma_1}^{\alpha}$ of the one-boundary Hilbert space is defined by considering all possible boundary conditions J that can be cut into two pieces by using a cut of the form $\Sigma = \Sigma_1$, so that the cut consists only of a single point. It is clear that any such boundary condition must contain an (oriented) interval $J = I \rightarrow K$ for some I, K , which is then cut by the choice of a point Σ_1 into the pieces $\psi_I := I \rightarrow$ and $\psi_K^* := \rightarrow K$. The ψ_I define a basis for a linear space V that we may call the space of allowed ‘Euclidean past’ boundary conditions. Similarly, the ψ_K^* define a basis for a linear space V^* that we may call the space of allowed ‘Euclidean future’ boundary conditions. We also define an anti-linear map $*$: $V \rightarrow V^*$ such that $(\psi_I)^* := \psi_I^*$, and a corresponding map $*$: $V^* \rightarrow V$ such that $(\psi_I^*)^* := \psi_I$.

The eigenvalues $(\psi_J, \psi_I)_{\alpha}$ then define a Hermitian inner product on V . Since this inner product was derived taking into account non-perturbative contributions associated with the sum-over-topologies, we will call it the non-perturbative inner product on V . Below, we will use $|I\rangle$ to denote the state $\psi_I \in V$, and we will write the non-perturbative inner product of such states in the form

$$\langle I|J\rangle_{\text{np}} := (\psi_I, \psi_J)_{\alpha}. \quad (3.1)$$

Taking the quotient of V by null vectors then yields the Hilbert space $\mathcal{H}_{\Sigma_1}^{\alpha}$. Note that this Hilbert space is automatically complete since V has finite dimension k . It is easy to see from (2.16) that, with probability one, the right-hand-side of (2.16) has rank $\min(k, d)$. It is thus clear that $\min(k, d)$ is the dimension of $\mathcal{H}_{\Sigma_1}^{\alpha}$.

In particular, equation (2.15) defines an ensemble of such Hilbert spaces $\mathcal{H}_{\Sigma_1}^{\alpha}$ with different dimensions d . Below, we will imagine that we have already selected a value of $\alpha = (d, \tilde{\alpha})$ from this ensemble. We will then refer to the associated $\mathcal{H}_{\Sigma_1}^{\alpha}$ as the non-perturbative Hilbert space $\mathcal{H}_{\text{np}}$.

The goal of our study below will be to compare the results of non-perturbative computations with those that would be obtained using a simpler perturbative description that does not include sums-over-topologies. Such perturbative computations cannot depend on α , as the α -parameters were already a result of non-perturbative effects from spacetime wormholes. We thus define the perturbative computation to be performed by taking the baby

universes to be in the no-boundary state $|\text{HH}\rangle$. With this understanding, we also take the perturbative results to be defined by using only the topology that would give the leading contribution at large S_0 . In particular, there is a perturbative Hermitian inner product $\langle I|J\rangle_{\text{p}}$ defined on V which we take to be given by⁴ e^{-2S_0} times the disk path integral with asymptotic boundary condition (ψ_I, ψ_J) . In particular, we define a conveniently normalized version $\mathcal{D}$ of the disk amplitude by the relations

$$\mathcal{D}[Z] = 1, \quad \text{and} \quad \mathcal{D}[(\psi_I, \psi_J)] := e^{-2S_0} \langle (\psi_I, \psi_J) \rangle_{\text{disk}} = \delta_{IJ} =: \langle I|J\rangle_{\text{p}}; \quad (3.2)$$

see again figure 3. Since this inner product is non-degenerate, it promotes the original linear space of Euclidean past boundary conditions V to a (perturbative) Hilbert space $\mathcal{H}_{\text{p}}$.

More generally, when the boundary condition has n connected components we similarly define $\mathcal{D}$ to be e^{-2nS_0} times the leading (n -disk) contribution to the associated path integral. With this definition, $\mathcal{D}$ factorizes over disconnected boundaries; e.g.

$$\mathcal{D}[(\psi_I, \psi_J)(\psi_K, \psi_L)] = \mathcal{D}[(\psi_I, \psi_J)] \mathcal{D}[(\psi_K, \psi_L)]. \quad (3.3)$$

For each fixed value of α , the non-perturbative inner product defines a self-adjoint operator M of rank $\min(k, d)$ on the perturbative Hilbert space with

$$\langle I|M|J\rangle_{\text{p}} := \langle I|J\rangle_{\text{np}} = (\psi_I, \psi_J)_{\alpha}. \quad (3.4)$$

In particular, although our notation does not indicate this explicitly, we emphasize that M is a different operator for each value of α . Using an overline to denote the average over the Wishart ensemble (2.16) with a given fixed value of d , we find

$$\overline{\langle I|M|J\rangle_{\text{p}}} = \delta_{IJ}. \quad (3.5)$$

As a result, the perturbative and non-perturbative inner products coincide *on average*, though they may be very different in particular elements of the ensemble. In particular, we know that the rank of M is $\min(k, d)$ while the rank of the perturbative inner product is k . Nevertheless, since the Wishart distribution is Gaussian, it is clear from (2.16) that fluctuations in given matrix elements $\langle I|M|J\rangle_{\text{p}}$ will be small whenever d is large.

3.2 Formalism and notation

Recall that the non-perturbative Hilbert space $\mathcal{H}_{\text{np}}$ is defined by taking the quotient of V with respect to the space of states of vanishing physical norm; i.e., $\mathcal{H}_{\text{np}} = V/\text{Ker } M$. Since the perturbative inner product was non-degenerate, we may write $\mathcal{H}_{\text{p}} = V$. Calling the above quotient map η , we may write

$$\eta : \mathcal{H}_{\text{p}} \rightarrow \mathcal{H}_{\text{np}}, \quad |\gamma\rangle \mapsto |\tilde{\gamma}\rangle, \quad (3.6)$$

where $|\tilde{\gamma}\rangle$ is the equivalence class in $V/\text{Ker } M$ of the state $|\gamma\rangle \in V$. We may alternatively think of η as an isomorphism from the quotient $\mathcal{H}_{\text{p}}/\text{Ker } M$ to $\mathcal{H}_{\text{np}}$.

⁴Including this factor of e^{-2S_0} gives a convenient overall normalization, though the results would be equivalent if one chose not to include this factor.

Note that since the eigenvalues of M are determined by the Wishart distribution, the map η is generally not an isometry even when acting on states in $\mathcal{H}_p$ orthogonal to $\text{Ker } M$. However, it will also be useful to isometrically embed the non-perturbative Hilbert space into $\mathcal{H}_p$. In order to do so, let us recall that, since M is Hermitian and positive semidefinite, there is a well-defined positive square root that we may denote by $X : \mathcal{H}_p \rightarrow \mathcal{H}_p$. As a result, if we define a map $\Upsilon : \mathcal{H}_{np} \rightarrow \mathcal{H}_p$ by

$$\Upsilon|\tilde{\gamma}\rangle := X|\gamma\rangle, \quad (3.7)$$

then Υ is an isometry. In particular, the perturbative inner product of $\Upsilon|\tilde{\gamma}\rangle$ and $\Upsilon|\tilde{\beta}\rangle$ is $\langle\gamma|\Upsilon^\dagger\Upsilon|\beta\rangle_p = \langle\gamma|X^\dagger X|\beta\rangle_p$, which agrees with $\langle\tilde{\gamma}|\tilde{\beta}\rangle_{np}$ by (3.4). Here and below, it will often be useful for clarity to decorate some X 's with daggers (†) even though X is self-adjoint (so that $X^\dagger = X$).

It will be useful to note that, as for any isometry, we have the relations $\Upsilon^\dagger\Upsilon = \mathbb{1}_p$ and $\Upsilon\Upsilon^\dagger = P_\Upsilon$, where P_Υ is the projection onto the range of Υ . In particular, we have $P_\Upsilon X = X$. We also comment that it is natural to think of Υ as defined by the ψ_I^a of the Wishart distribution in the sense that there is another (not orthonormal, possibly overcomplete) basis $|\hat{a}\rangle$ of $\mathcal{H}_{np}$ such that $\Upsilon|\hat{a}\rangle = \sum_I \psi_I^a |I\rangle$. Below, we will often use the isometry Υ to identify a state $|\tilde{\alpha}\rangle$ in the quotient $V/\text{Ker } M$ with $X|\alpha\rangle \in XV \subset V$. Here XV is the range of the linear operator X acting on the linear space V .

Let us now make two further brief observations. The first is that $X = \Upsilon \circ \eta$. The second is that if, for every element of our ensemble, X were of the form $\mathcal{N}P$ for a random projection P and some fixed normalization constant $\mathcal{N}$, then for $k > d$ the condition (3.5) and the fact that X has rank $\min(k, d)$ would imply $\mathcal{N} = \sqrt{k/d}$. We will show in appendix A.3 that X is indeed well-approximated by an operator of this form in the limit $k \gg d$.

3.3 Operator insertions and time evolution

The above inner products are computed by path integrals with a single boundary (ψ_I, ψ_J) . We will be interested in path integrals that describe the further insertion of operators at that boundary. Recall, however, that the allowed boundary conditions in the model are just unions of the circles (Z 's) and labelled intervals (ψ_I, ψ_J) . Suppose then that we wish to insert the perturbative operator $|K\rangle_p \langle L|$, where the notation ${}_p \langle L|$ indicates that this dual-state acts on vectors in V in the manner defined by the perturbative inner product.⁵ We will then consider boundaries that contain two labelled intervals, (ψ_I, ψ_K) and (ψ_L, ψ_J) which yields (3.3); see figure 4. Of course, a general operator $\mathcal{O}$ on $\mathcal{H}_p$ can be written in the form

$$\mathcal{O} = \sum_{K,L} \mathcal{O}_{K,L} |K\rangle_p \langle L|. \quad (3.8)$$

In fact, for any Hilbert space $\mathcal{H}_{\text{ext}}$, we can write a general operator on $\mathcal{H}_p \otimes \mathcal{H}_{\text{ext}}$ in this form by taking the coefficients $\mathcal{O}_{K,L}$ to be operators on $\mathcal{H}_{\text{ext}}$. Let us now introduce the

⁵We now see that it would have been better to denote the perturbative inner product of $|I\rangle, |J\rangle \in V$ with a left-subscript (${}_p \langle I|J\rangle$) rather than a right-subscript ($\langle I|J\rangle_p$), though we will continue to use the right-subscript as this notation will be more familiar to many readers.

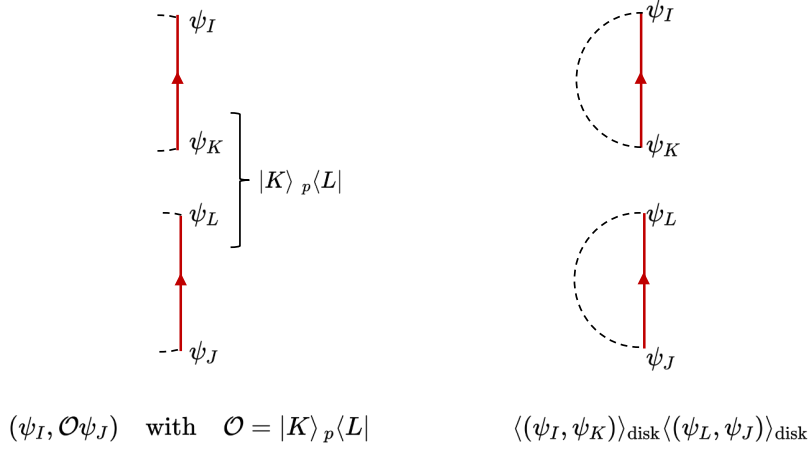


Figure 4. The two labelled intervals (ψ_I, ψ_K) and (ψ_L, ψ_J) (left) that result from the insertion of the operator $|K\rangle_p \langle L|$ into (ψ_I, ψ_K) , and the disk contribution to the path integral with this boundary condition (right).

symbol ζ_α to denote the full non-perturbative path integral in a given baby universe α -state. Applying ζ_α to the above boundary condition gives

$$\begin{aligned}
 \zeta_\alpha \left[\sum_{K,L} \mathcal{O}_{K,L}(\psi_I, \psi_K)(\psi_L, \psi_J) \right] &= \sum_{K,L} \mathcal{O}_{K,L} \langle I|K \rangle_{\text{np}} \langle L|J \rangle_{\text{np}} \\
 &= \sum_{K,L} \mathcal{O}_{K,L} \langle I|M|K \rangle_p \langle L|M|J \rangle_p \\
 &= \langle I|X^\dagger (X\mathcal{O}X^\dagger) X|J \rangle_p = \langle \tilde{I}|\Upsilon^\dagger (X\mathcal{O}X^\dagger) \Upsilon|\tilde{J} \rangle_{\text{np}}. \quad (3.9)
 \end{aligned}$$

In the last equality we have used the definition (3.7) and the immediate consequence

$$\begin{aligned}
 \langle I|X|J \rangle_p &= \langle J|X|I \rangle_p^* = \langle J|\Upsilon\tilde{I} \rangle_p^* \\
 &= \langle \Upsilon^\dagger J|\tilde{I} \rangle_{\text{np}}^* = \langle \tilde{I}|\Upsilon^\dagger J \rangle_{\text{np}}, \quad (3.10)
 \end{aligned}$$

where $|\Upsilon\tilde{I}\rangle := \Upsilon|\tilde{I}\rangle \in XV$ and $|\Upsilon^\dagger J\rangle := \Upsilon^\dagger|J\rangle \in V/\text{Ker } M$. The result (3.9) suggests that we should associate any operator $\mathcal{O}$ on $\mathcal{H}_p$ with a gravitationally dressed operator $\tilde{\mathcal{O}}$ on the non-perturbative Hilbert space of the form

$$\tilde{\mathcal{O}} = \Upsilon^\dagger X\mathcal{O}X^\dagger \Upsilon. \quad (3.11)$$

Indeed, if we write down a corresponding path integral that, at the disk level, would compute the correlator $\langle I|\mathcal{O}_1\mathcal{O}_2\ldots\mathcal{O}_n|J \rangle_p$, then at the non-perturbative level we instead find the result $\langle \tilde{I}|\tilde{\mathcal{O}}_1\tilde{\mathcal{O}}_2\ldots\tilde{\mathcal{O}}_n|\tilde{J} \rangle_{\text{np}}$. Henceforth, it will be convenient to use Υ to identify $\mathcal{H}_{\text{np}}$ with its isometric embedding in $\mathcal{H}_p$ as previously advertised. In doing so, we will write (3.11) in the simpler form

$$\tilde{\mathcal{O}} = X\mathcal{O}X^\dagger. \quad (3.12)$$

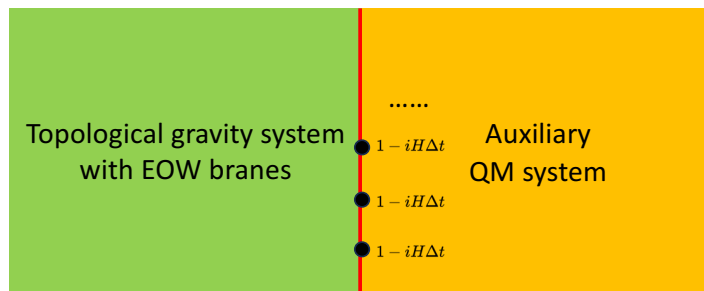


Figure 5. Boundary insertions can be used to couple our topological model to an external non-gravitational ‘bath’ system.

In the case where the $\mathcal{O}_{K,L}$ are operators on some external Hilbert space $\mathcal{H}_{\text{ext}}$, it would be more explicit to write (3.12) in the form

$$\tilde{\mathcal{O}} = (X \otimes \mathbb{1}_{\text{ext}}) \mathcal{O} (X^\dagger \otimes \mathbb{1}_{\text{ext}}), \quad (3.13)$$

where $\mathbb{1}_{\text{ext}}$ is the identity on $\mathcal{H}_{\text{ext}}$.

Although the original topological model has no evolution (i.e., its Hamiltonian H_0 vanishes identically), we can now use this formalism to introduce and study an arbitrary notion of time-evolution for this model. In particular, let us consider a one-parameter family of operators $H(t)$ on the perturbative one-boundary Hilbert space $\mathcal{H}_p$. Let us then choose times t_n for all $n \in \mathbb{Z}$ such that $t_n \rightarrow \pm\infty$ as $n \rightarrow \pm\infty$ and use the representation (3.8) to define boundary conditions corresponding to the insertion of the sequence of operators $1 - i(t_n - t_{n-1})H(t_n)$. In the limit where we increase the density of times t_n such that $t_n - t_{n-1} \rightarrow 0$ for all n , the corresponding (rescaled) disk path integral $\mathcal{D}$ computes matrix elements of the path-ordered-exponential operator $\mathcal{P} \exp(-\int iH(t)dt)$. It thus implements the time evolution associated with $H(t)$. As noted above, this can also include the introduction of interactions between our topological model and an external system $\mathcal{H}_{\text{ext}}$; see figure 5.

As in figure 4, each insertion of an operator of the form (3.8) increases the number of connected components of the boundary by one. Since each component computes an inner product, and since the non-perturbative inner product inserts an X (or $X^\dagger$) next to each bra or ket, we see immediately that the non-perturbative path integral with such insertions describes evolution under the gravitationally dressed Hamiltonian

$$\tilde{H}(t) = XH(t)X^\dagger. \quad (3.14)$$

Furthermore, it is clear from (3.14) that the dressed Hamiltonian is self-adjoint, so that time evolution on the non-perturbative Hilbert space is unitary. In contrast, a less careful analysis might instead lead one to believe that one could use interactions with $\mathcal{H}_{\text{ext}}$ to evolve our gravitating system from a non-trivial state to a null state (which would be a clear violation of unitarity). Indeed, one can easily write down scenarios where the *perturbative* evolution under $\mathcal{P} \exp(-\int iH(t)dt)$ would transform our system from a non-trivial state $|\psi\rangle$ (for which $\eta|\psi\rangle \neq 0$) into a null state annihilated by η . However, we also see that non-perturbative corrections will then necessarily modify the dynamics such that the gravitationally dressed evolution $\mathcal{P} \exp(-\int i\tilde{H}(t)dt)$ remains unitary. We will discuss this issue again in section 6.

4 Large discrepancies for generic insertions

Having introduced both our general formalism and the specific toy model, we now turn to the question of how the perturbative and non-perturbative time-evolutions compare for path integrals of the form described in section 3.3; see again figure 5. In performing such analysis we will often compute averaged quantities in order to derive results applicable to typical α sectors, but the reader should keep in mind that the dynamics we define will act within each α sector separately.

We consider the case $d < k$ so that the dimensions of $\mathcal{H}_p$ and $\mathcal{H}_{np}$ differ, and we will be most interested in the case $d \ll k$. We may take the initial and final states to be described by the same boundary condition so that we study only expectation values. This is not actually a restriction since matrix elements of an operator between states $|\phi_1\rangle$ and $|\phi_2\rangle$ can always be expressed as linear combinations of expectation values of the operator in states of the form $c_1|\phi_1\rangle + c_2|\phi_2\rangle$. For simplicity, we will also take $H(t)$ to be independent of time. We thus wish to choose $|\psi\rangle \in \mathcal{H}_p$ with $|\tilde{\psi}\rangle := \eta|\psi\rangle \in \mathcal{H}_{np}$ and to compare

$$\langle\psi|e^{-iHt}|\psi\rangle_p \quad (4.1)$$

with

$$\langle\tilde{\psi}|e^{-i\tilde{H}t}|\tilde{\psi}\rangle_{np} = \langle\psi|X^\dagger e^{-i(XHX^\dagger)t}X|\psi\rangle_p. \quad (4.2)$$

We will shortly find that, even for short times t of order $\sqrt{d/k}$ for $d \ll k$, the evolutions (4.1) and (4.2) are starkly different for general choices of H and $|\psi\rangle$. One way to see this is to compute averages of traces of powers of $\tilde{H}$ over the Wishart ensemble (again, for a fixed choice of d). Note that, if we choose an orthonormal basis $|\tilde{I}\rangle$ of $\mathcal{H}_{np}$ for $I = 1, \dots, d$, then each state $|\tilde{I}\rangle$ must be of the form $\eta|I\rangle$ for some $|I\rangle \in V$, and our isometric embedding Υ (3.7) then tells us that the trace of any operator $\tilde{O}$ on $\mathcal{H}_{np}$ can be written in the form

$$\text{Tr}_{np} \tilde{O} := \sum_I \langle\tilde{I}|\tilde{O}|\tilde{I}\rangle_{np} = \sum_I \langle I|X^\dagger \Upsilon \tilde{O} \Upsilon^\dagger X|I\rangle_p, \quad (4.3)$$

where we recall that $\Upsilon \Upsilon^\dagger = P_\Upsilon$ (where P_Υ is the projection onto the range of Υ , which is also the range of X) and that $\Upsilon^\dagger \Upsilon = \mathbb{1}_{np}$, so that $\Upsilon^\dagger$ is the left-inverse of Υ . In particular, for $\tilde{O} = \tilde{H}^n = [\Upsilon^\dagger X H X^\dagger \Upsilon]^n = \Upsilon^\dagger [X H X] X^n \Upsilon$, so that the result (4.3) gives

$$\text{Tr}_{np}(\tilde{H}^n) = \sum_{I=1}^d \langle I|X^\dagger [X H X^\dagger]^n X|I\rangle_p = \text{Tr}_p([X H X^\dagger]^n) = \text{Tr}_p([M H]^n), \quad (4.4)$$

where Tr_{np} denotes the trace on the non-perturbative Hilbert space $\mathcal{H}_{np}$, and Tr_p is the trace on the perturbative Hilbert space $\mathcal{H}_p$. The second equality in (4.4) follows from the fact that, since $|\tilde{I}\rangle_{np}$ is an orthonormal basis and Υ is an isometry, the states $\Upsilon|\tilde{I}\rangle = X|I\rangle$ form an orthonormal basis for the orthogonal complement of the kernel of X .

Recall also that the Wishart ensemble is built from Gaussian variables. As a result, so long as one can write the desired expression in terms of M rather than just X , ensemble expectation values are straightforward to compute by performing appropriate Wick contractions; see appendix A. In particular, using (3.5) immediately yields

$$\overline{\text{Tr}_{np} \tilde{H}} = \text{Tr}_p H. \quad (4.5)$$

The result (4.5) is already interesting since H has k eigenvalues while $\tilde{H}$ has only d eigenvalues. As a result, typical eigenvalues of $\tilde{H}$ must be larger than those of H by a factor of k/d .

The result (4.5) leaves open the possibility that the discrepancy might just be an overall additive shift in the eigenvalues, which would then give only an overall phase in the time evolution. However, this possibility is easily ruled out by considering the operator $\Delta := \tilde{H} - c\mathbb{1}_{\text{np}}$ with $c = \frac{1}{d} \text{Tr}_{\text{p}} H$, so that $\overline{\text{Tr}_{\text{np}} \Delta} = 0$. Using cyclicity of the trace together with (A.10) and (A.11) yields

$$\overline{\text{Tr}_{\text{np}}(\Delta^2)} = \overline{\text{Tr}_{\text{p}}(XHX^\dagger XHX^\dagger - 2cXHX^\dagger + c^2P_{\text{r}})} = \text{Tr}_{\text{p}}(H^2). \quad (4.6)$$

Again, the trace on the left hand side of (4.6) is performed over a space of dimension d , while the trace on the right is over a space of dimension k . As a result, if there is a meaning to discussing typical eigenvalues of Δ , they must have absolute values that are roughly $\sqrt{k/d}$ times the absolute values of typical eigenvalues of H . Since $\sqrt{k/d}$ is large and $\tilde{H} = c\mathbb{1}_{\text{np}} + \Delta$, even if we ignore the overall phase the evolution generated by $\tilde{H}$ will still be more rapid than that generated by the perturbative Hamiltonian H . (However, when $\text{Tr}_{\text{p}} H$ is dominated by its positive eigenvalues it is also true that $\tilde{H}$ is close to a multiple of the identity in the sense that c is large compared with the magnitudes of the typical eigenvalues of Δ .)

We have also computed the next two moments of Δ . Performing the relevant Wick contractions yields

$$\overline{\text{Tr}_{\text{np}}(\Delta^3)} = \frac{d^2 + 1}{d^2} \text{Tr}_{\text{p}}(H^3), \quad (4.7)$$

and

$$\overline{\text{Tr}_{\text{np}}(\Delta^4)} = \left(\frac{1}{d} + \frac{5}{d^2}\right) \text{Tr}_{\text{p}}(H^4) + \left(\frac{2}{d} + \frac{1}{d^3}\right) \left(\text{Tr}_{\text{p}}(H^2)\right)^2. \quad (4.8)$$

These results are again consistent with typical eigenvalues of Δ having absolute values that are roughly $\sqrt{k/d}$ times the absolute values of typical eigenvalues of H . The result (4.7) also suggests that, at leading order in k/d , the eigenvalues of Δ are distributed evenly on both sides of zero, but that there is a subleading asymmetry dictated by traces of odd powers of H .

Given that the $d \times d$ matrix $\tilde{H}$ is constructed from the $k \times k$ matrix H by the random compression $\tilde{H} = XHX^\dagger$, such large discrepancies should not be a surprise. Indeed, since the ensemble average over $\tilde{H}$ is invariant under all unitary transformations on $\mathcal{H}_{\text{p}}$, and since we expect fluctuations to be small for $k \gg d$, it is natural to expect that $\tilde{H}$ is proportional to the identity at leading order in k/d , and that the remaining part of $\tilde{H}$ given by Δ should consist largely of random junk that has little to do with the detailed dynamics of H (though the distribution of eigenvalues will be influenced by the corresponding distribution for H). Indeed, if instead the eigenvalues of Δ^2 became sharply peaked in the limit of large d and large k/d , then the coefficient of the $(\text{Tr}_{\text{p}}(H^2))^2$ term on the right-hand-side of (4.8) would approach $1/d$ at large d . The fact that it instead approaches $2/d$ thus supports the idea that the eigenvalues of Δ are rather random.

Further support for this picture is provided by considering simple special cases with small d . For example, if $d = 1$, then X picks out a single random vector in $\mathcal{H}_{\text{p}}$ to represent the one non-trivial eigenstate of $\mathcal{H}_{\text{np}}$. Since the direction of this vector is chosen uniformly, up to fluctuations that are small at large k it must consist of roughly equal components along

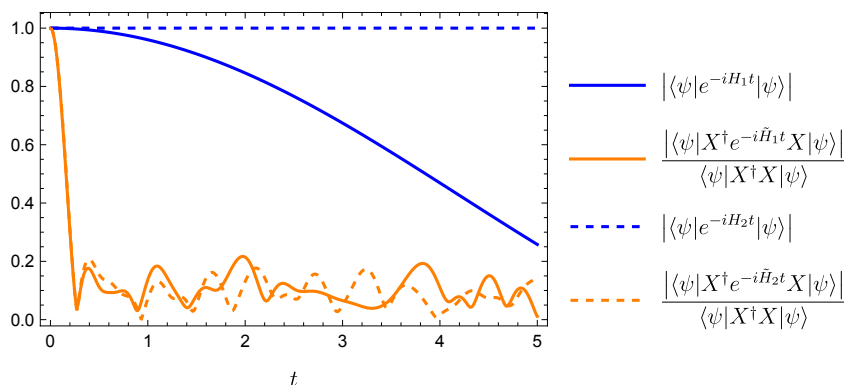


Figure 6. Comparison of two perturbative time evolutions (4.1) (blue solid and dashed lines) with the corresponding non-perturbative time evolutions (4.2) (gold solid and dashed lines) for a fixed random draw from the Wishart ensemble with $k = D = 2000$ and $d = 100$ for a randomly chosen state $|\psi\rangle \in \mathcal{H}_p$. The perturbative Hamiltonian H_1 was chosen to have a uniformly spaced spectrum with eigenvalues between $E_{\min} = 1$ and $E_{\max} = 2$. In contrast, H_2 is proportional to identity but was normalized to have the same trace as H_1 . While the details of such plots depend on the random draw from the ensemble, the qualitative features shown above are typical.

each eigenvector of H . Thus $XHX^\dagger$ is proportional to P_Γ with a coefficient proportional to $\text{Tr}_p(H)$. Similarly, for $d = 2$, the Wishart distribution will build X from a pair of random states. Such states are nearly orthogonal at large k , and both will again have roughly equal components along each eigenvector of H . So both eigenvalues of H will be roughly equal and will differ only by random fluctuations. This will then continue to be the case for all $d \ll k$.

As final confirmation of this picture, figure 6 shows numerical results comparing two perturbative time evolutions (4.1) (blue solid and dashed lines) with the corresponding non-perturbative time evolutions (4.2) (gold solid and dashed lines) for a fixed random draw from the Wishart ensemble⁶ with $k = D = 2000$ (where D is the rank of H) and $d = 100$ for a randomly chosen state $|\psi\rangle \in \mathcal{H}_p$. The perturbative Hamiltonian H_1 was chosen to have a uniformly spaced spectrum with eigenvalues between $E_{\min} = 1$ and $E_{\max} = 2$. In contrast, H_2 is proportional to the identity but was normalized to have the same trace as H_1 . While the associated perturbative evolutions are very different, the initial parts of the corresponding non-perturbative time evolutions in figure 6 are nearly identical. We interpret this as being due to the fact that (4.6) implies typical eigenvalues of Δ for H_1 and H_2 to be related by a factor of $\sqrt{\text{Tr}_p(H_1^2)/\text{Tr}_p(H_2^2)} \sim 1.018$; i.e., they differ only at the 2% level. The higher moments will be closely related as well.

Choosing special states $|\psi\rangle$ instead of random states does not appear to improve the situation. In particular, if $|E_{\max}\rangle$ is the eigenstate of H_1 with highest eigenvalue, taking either $|\psi\rangle = |E_{\max}\rangle$ or $|\psi\rangle = X|E_{\max}\rangle$ yields plots that look qualitatively similar to those already shown in figure 6.

⁶The built-in Mathematica function `WishartMatrixDistribution` is defined only for $d \geq k$. In contrast, we are interested in the $k > d$ case. The matrices used to generate our numerical results were thus built using the Mathematical function `NormalDistribution` and equation (A.3).

5 Long-time agreement for small-rank insertions

Having illustrated the stark difference between perturbative and non-perturbative time evolutions for generic H in section 4, it is natural to ask whether there is better agreement for certain choices of H . One context where this is to be expected occurs when the rank D of H is small but the dimension d of $\mathcal{H}_{\text{np}}$ is large. At a heuristic level, this follows from the fact that, for a collection of orthonormal states $|I\rangle \in \mathcal{H}_{\text{p}}$ (with $I = 1, \dots, D$), when $D \ll d$, with high probability one finds the corresponding non-perturbative states $|\tilde{I}\rangle := \eta|I\rangle$ to be approximately orthonormal. As a result, if $H = \sum_{A=1}^D \lambda_A |\phi_A\rangle\langle\phi_A|$, then each $X|\phi_A\rangle$ is an approximate eigenvector of $XHX^\dagger$ with eigenvalue λ_A . This means that the dynamics generated by H on the subspace spanned by the $|\phi_A\rangle$ will approximate that generated by $\tilde{H}$ on the subspace spanned by the images $\eta|\phi_A\rangle$. Furthermore, since $d \gg D$, a state $|\phi_\perp\rangle$ annihilated by H , and thus orthogonal to all $|\phi_A\rangle$, will have $\eta|\phi_\perp\rangle$ approximately orthogonal to all $\eta|\phi_A\rangle$ and will thus be approximately annihilated by $\tilde{H}$. Thus the perturbative and non-perturbative path integrals should generate the same dynamics. The rest of this section will make the above argument precise and will establish quantitative estimates of the associated errors.

We will also choose to make a stronger comparison between the perturbative and non-perturbative time evolutions than in the above sections. Rather than merely comparing matrix elements of the time evolution operators between corresponding pairs of states, we will now directly compare the final states obtained after time-evolving an initial state. Specifically, after choosing a state $|\psi\rangle \in \mathcal{H}_{\text{p}}$, we will compare the non-perturbative states $e^{i\tilde{H}t}\eta|\psi\rangle$ and $\eta e^{iHt}|\psi\rangle$. Since for typical states $|\phi_1\rangle, |\phi_2\rangle$ the perturbative inner product $\langle\phi_1|\phi_2\rangle_{\text{p}}$ is (with high probability) a good approximation to the non-perturbative inner product of $\eta|\phi_1\rangle$ and $\eta|\phi_2\rangle$, close agreement between $e^{i\tilde{H}t}\eta|\psi\rangle$ and $\eta e^{iHt}|\psi\rangle$ will imply (again, with high probability) close agreement between (4.1) and (4.2). It is in this sense that the comparison below will constitute a more stringent test of agreement than that used in section 4 above. We also mention that $\eta e^{iHt}|\psi\rangle$ is the quantity that appears when we can approximate the full set of non-perturbative effects as the insertion of a ‘black hole final state’ (postselection) condition [16]. It is thus clear that this approximation will also fail drastically in the large-rank context that was studied above in section 4.

Below, it will often be useful to directly discuss the corresponding operator

$$E_1(t) := \eta e^{-iHt} = \Upsilon^\dagger X e^{-iHt}, \quad (5.1)$$

which enacts perturbative evolution followed by the action of η , as well as

$$E_2(t) := e^{-i\tilde{H}t}\eta = \Upsilon^\dagger e^{-iXHX^\dagger t}X, \quad (5.2)$$

which describes non-perturbative evolution following the action of η . In writing (5.1) and (5.2) we have used the relation $\Upsilon^\dagger X = \eta$ which follows from the identities $\Upsilon^\dagger \Upsilon = \mathbb{1}_{\text{np}}$ and $X = \Upsilon \eta$.

Note that $E_1(t)$ and $E_2(t)$ are both maps from $\mathcal{H}_{\text{p}}$ to $\mathcal{H}_{\text{np}}$. As above, we will generally use the isometry Υ to identify $\mathcal{H}_{\text{np}}$ with its image $\Upsilon\mathcal{H}_{\text{np}}$ and we will thus omit the factors of $\Upsilon^\dagger$ from E_1 and E_2 . We will also introduce a notion of ‘generic’ states in $\mathcal{H}_{\text{p}}$, meaning that such states are chosen without knowledge of the particular matrix $M = X^\dagger X$ drawn from the Wishart distribution. We will see below that for $D \ll d$, even for long times t of

order d^α with $\alpha < \frac{1}{2}$, with high probability the operators $E_1(t), E_2(t)$ will agree well when the eigenvalues of H are well-separated and when these operators act on generic states.

5.1 Eigenvalues and eigenvectors

In order to compare (5.1) and (5.2), it will be useful to first develop some understanding of at least approximate eigenvalues and eigenvectors for the non-perturbative Hamiltonian $\tilde{H} = XHX^\dagger$.

Recall from (3.5) that the ensemble average of the physical inner product is

$$\overline{\langle I|X^\dagger X|J\rangle} = \delta_{IJ}. \quad (5.3)$$

We can also use (A.10) to find

$$\overline{|\langle I|X^\dagger X|J\rangle|^2} = \overline{M_{IJ}M_{IJ}^*} = \delta_{IJ} + \frac{1}{d}, \quad (5.4)$$

where I, J are not summed. The variance of any given matrix element $\langle I|X^\dagger X|J\rangle$ is thus $1/d$ for all I, J .

Denoting the nontrivial eigenvalues of H by λ_A and the corresponding orthonormal eigenvectors by $|\phi_A\rangle$ with $A = 1, \dots, D$, we may write the perturbative Hamiltonian H in the form

$$H = \sum_{A=1}^D \lambda_A |\phi_A\rangle\langle\phi_A|. \quad (5.5)$$

We will assume that we can write (5.5) in a form that is independent of d and then study the limit $d \rightarrow \infty$. For the moment, we also assume the eigenvalues to be non-degenerate. We will consider the degenerate case separately in section 5.3. The non-perturbative Hamiltonian is correspondingly written

$$\tilde{H} \equiv XHX^\dagger = \sum_{A=1}^D \lambda_A X|\phi_A\rangle {}_p\langle\phi_A|X^\dagger, \quad (5.6)$$

where the notation ${}_p\langle\phi_A|$ again indicates the linear functional on $V = \mathcal{H}_p$ that acts on any $|\psi\rangle$ to give the perturbative inner product $\langle\phi_A|\psi\rangle_p$.

For $D < d$, the D states $X|\phi_A\rangle$ will be linearly independent with probability one (though they will not generally be orthogonal). This occurs because the states $|\phi_A\rangle$ are linearly independent, so that fine-tuning is required for the action of the map X to render them linearly dependent. Thus $XHX^\dagger$ has rank D with probability 1. However, only in the $D \ll d$ limit will the states $X|\phi_A\rangle$ become approximately orthonormal. In particular, let us consider applying the Gram-Schmidt orthonormalization procedure to the states $X|\phi_A\rangle$. So long as $|\phi_A\rangle, |\phi_B\rangle$ are generic states, from (A.6) and (A.10) we know that

$$\langle\phi_A|X^\dagger X|\phi_B\rangle = \delta_{AB} + \epsilon_{AB}, \quad (5.7)$$

where

$$\epsilon_{AB} = |\epsilon_{AB}|e^{i\theta_{AB}}, \quad (5.8)$$

$|\epsilon_{AB}| = O(1/\sqrt{d})$, and for $A \neq B$ the above θ_{AB} is a random phase.⁷ As a result, the D th vector generated by the Gram-Schmidt procedure will differ from $X|\phi_D\rangle$ by $D - 1$ approximately orthogonal terms of order $\sqrt{1/d}$, so that the magnitude of the difference from $X|\phi_D\rangle$ is of order $\sqrt{\frac{D}{d}}$. Thus for $D \ll d$ the states $\{X|\phi_A\rangle\}$ are orthonormal with probability 1. In this limit (holding fixed the non-trivial eigenvalues λ_A), we see from (5.6) that the eigenstates of $\tilde{H}$ become precisely $\{X|\phi_A\rangle\}$ with eigenvalues λ_A .

We can also expand $\tilde{H} = XHX$ in terms of its own nontrivial eigenvectors $|\tilde{\Phi}_A\rangle$ and the corresponding eigenvalues $\tilde{\lambda}_A$. Note that while $|\tilde{\Phi}_A\rangle$ must be $\eta|\Phi_A\rangle$ for some state $|\Phi_A\rangle \in V$, this $|\Phi_A\rangle$ will generally not be an eigenvector of the perturbative Hamiltonian H . This is the reason that eigenvectors of $\tilde{H}$ should not be called $|\tilde{\phi}_A\rangle$.

As argued above, $\tilde{H} = XHX^\dagger$ has rank D with probability 1, so we again take $A = 1, \dots, D$ and write

$$\begin{aligned} XHX^\dagger &= \Upsilon \tilde{H} \Upsilon^\dagger = \sum_{A=1}^D \tilde{\lambda}_A \Upsilon \frac{|\tilde{\Phi}_A\rangle_{\text{np}} \langle \tilde{\Phi}_A|}{\langle \tilde{\Phi}_A | \tilde{\Phi}_A \rangle_{\text{np}}} \Upsilon^\dagger \\ &= \sum_{A=1}^D \tilde{\lambda}_A \Upsilon \frac{\eta|\Phi_A\rangle_{\text{p}} \langle \Phi_A| \eta^\dagger}{\langle \Phi_A | X^\dagger X | \Phi_A \rangle_{\text{p}}} \Upsilon^\dagger \\ &= \sum_{A=1}^D \tilde{\lambda}_A \frac{X|\Phi_A\rangle_{\text{p}} \langle \Phi_A| X^\dagger}{\langle \Phi_A | X^\dagger X | \Phi_A \rangle_{\text{p}}}, \end{aligned} \quad (5.9)$$

where in the last step we have used $X = \Upsilon \circ \eta$ and thus $\eta^\dagger \circ \Upsilon^\dagger = X^\dagger$. The notation $_{\text{p}}\langle \Phi_A |$ again indicates the linear functional on $V = \mathcal{H}_{\text{p}}$ that acts on any $|\phi\rangle$ to give the perturbative inner product $\langle \Phi_A | \phi \rangle_{\text{p}}$. Note that we allow arbitrary normalizations for the $|\tilde{\Phi}_A\rangle$ because we would like to approximate $\Upsilon|\tilde{\Phi}_A\rangle$ by $X|\phi_A\rangle = \Upsilon\eta|\phi_A\rangle$. We also define $\tilde{P}_0$ to be the projection operator onto the kernel of $\tilde{H}$.

We now explore the details of how $\tilde{\lambda}_A, |\tilde{\Phi}_A\rangle$ approach $\lambda_A, X|\phi_A\rangle$ as $d \rightarrow \infty$. In order to do so, it is useful to note that while states in the kernel of H are orthogonal to those in the range of H , this will no longer be true of the images obtained by acting on the above states with the operator X . In particular, for an arbitrary normalized state $|\phi_0\rangle$, if the projection map to the kernel of H is P_0 , we find $\langle \phi_A | X^\dagger X P_0 | \phi_0 \rangle_{\text{p}} = O(1/\sqrt{d})$.

To avoid writing out this matrix element repeatedly below, we define the symbol $\varepsilon_A := \langle \phi_A | X^\dagger X P_0 | \phi_0 \rangle_{\text{p}} = O(1/\sqrt{d})$. Furthermore, from (A.6) and (A.10) we have

$$\overline{\langle \phi_A | X^\dagger X P_0 | \phi_0 \rangle_{\text{p}}} = 0, \quad \text{and} \quad |\overline{\langle \phi_A | X^\dagger X P_0 | \phi_0 \rangle_{\text{p}}}|^2 = \frac{||P_0|\phi_0\rangle||_{\text{p}}^2}{d}. \quad (5.10)$$

As we will see, the small parameters ϵ_{AB} and ε_A will control the discrepancy between the action of $E_1(t)$ and $E_2(t)$ on a generic state. We will compute this discrepancy to leading order in these parameters.

⁷However, Hermiticity of $X^\dagger X$ requires $\epsilon_{AA} \in \mathbb{R}$ and thus $\theta_{AA} = 0$ or $\theta_{AA} = \pi$. More generally, hermiticity of $X^\dagger X$ requires $\epsilon_{AB} = \epsilon_{BA}^*$.

We begin by studying the rate at which each non-perturbative eigenvalue $\{\tilde{\lambda}_A\}$ converges to $\{\lambda_A\}$. In particular, we use (3.14), (5.5), and (5.7) to write

$$\tilde{H}X|\phi_B\rangle = XHX^\dagger X|\phi_B\rangle = \sum_A \lambda_A X|\phi_A\rangle \langle \phi_A|X^\dagger X|\phi_B\rangle_p = \lambda_B X|\phi_B\rangle + O\left(\sqrt{\frac{D}{d}}\right). \quad (5.11)$$

To see that the error term is of order $\sqrt{\frac{D}{d}}$, one may calculate the norm of the difference between the left side and the explicit term on the right. This yields

$$\left\|\left(\tilde{H}X - \lambda_B X\right)|\phi_B\rangle\right\|^2 = \frac{\text{Tr}_p(H^2)}{d} + \frac{\lambda_B^2 + (\text{Tr}_p H)^2}{d^2}. \quad (5.12)$$

Since $\text{Tr}_p(H^2), \text{Tr}_p(H) = O(D)$, it follows that the norm squared is of order $\frac{D}{d}$, hence the error term in (5.11).

While this result already shows $\tilde{\lambda}_A \rightarrow \lambda_A$ and $X|\phi_A\rangle \rightarrow |\tilde{\Phi}_A\rangle$ as $d/D \rightarrow \infty$, the error terms can be characterized more explicitly by introducing the quantities

$$\delta\lambda_A \equiv \tilde{\lambda}_A - \lambda_A \quad \text{and} \quad |\Delta_A\rangle \equiv |\tilde{\Phi}_A\rangle - X|\phi_A\rangle. \quad (5.13)$$

The normalization of $|\tilde{\Phi}_A\rangle$ will be chosen to set $\langle \Delta_A|X|\phi_A\rangle_p = 0$. We then have $\delta\lambda_A = O(1/\sqrt{d})$, $|\Delta_A\rangle = O(\sqrt{\frac{D}{d}})$, where the notation $|\Delta_A\rangle = O(\sqrt{\frac{D}{d}})$ means $\| |\Delta_A\rangle \|_p = O(\sqrt{\frac{D}{d}})$; i.e., $\langle \Delta_A|\Delta_A\rangle_p = O(D/d)$. Thus we find

$$\begin{aligned} \tilde{\lambda}_A |\tilde{\Phi}_A\rangle &= \tilde{H}|\tilde{\Phi}_A\rangle = XHX^\dagger |\tilde{\Phi}_A\rangle \\ &= \sum_B \lambda_B X|\phi_B\rangle_p \langle \phi_B|X^\dagger (X|\phi_A\rangle + |\Delta_A\rangle) \\ &= \lambda_A X|\phi_A\rangle + \sum_B \lambda_B X|\phi_B\rangle \epsilon_{BA} + \sum_B \lambda_B X|\phi_B\rangle \langle \phi_B|X^\dagger |\Delta_A\rangle_p, \end{aligned} \quad (5.14)$$

where in the last step we have again used equation (5.7). At the same time, we have

$$\tilde{\lambda}_A |\tilde{\Phi}_A\rangle = (\lambda_A + \delta\lambda_A)(X|\phi_A\rangle + |\Delta_A\rangle) = \lambda_A X|\phi_A\rangle + \delta\lambda_A X|\phi_A\rangle + \lambda_A |\Delta_A\rangle + O(\sqrt{D}/d). \quad (5.15)$$

Putting these together yields

$$\delta\lambda_A X|\phi_A\rangle + \lambda_A |\Delta_A\rangle = \sum_B \lambda_B X|\phi_B\rangle \epsilon_{BA} + \sum_B \lambda_B X|\phi_B\rangle \langle \phi_B|X^\dagger |\Delta_A\rangle_p + O(\sqrt{D}/d). \quad (5.16)$$

Taking the inner product with $X|\phi_A\rangle$ and keeping only first-order corrections we obtain

$$\delta\lambda_A = \lambda_A \epsilon_{AA} + O(D/d). \quad (5.17)$$

In computing the above error term, we have noted that the range of $XHX^\dagger$ is spanned by the states $X|\phi_B\rangle$ so that, since we defined $|\Delta_A\rangle$ to be orthogonal to $X|\phi_A\rangle$, $|\Delta_A\rangle$ is in fact a superposition of the $D - 1$ states $X|\phi_B\rangle$ for $B \neq A$. We have then made the natural assumption that the orders of each matrix element $\langle \phi_B|X^\dagger |\Delta_A\rangle_p$ in both d and D are independent of B , and thus that they are of order $(D^0/\sqrt{d})$. This assumption will be justified below. We also used the following relation

$$\sum_B \lambda_B \langle \phi_A|X^\dagger X|\phi_B\rangle \epsilon_{BA} = \lambda_A \epsilon_{AA} + \sum_B \lambda_B |\epsilon_{AB}|^2, \quad (5.18)$$

which follows from (5.7). If we now insert (5.17) into equation (5.16), take the inner product with $X|\phi_B\rangle$, and keep only first-order corrections we find

$$\langle \Delta_A | X | \phi_B \rangle_p = \frac{\lambda_B}{\lambda_A - \lambda_B} \epsilon_{AB} (1 - \delta_{AB}) + O(D/d), \quad (5.19)$$

whence (5.16) yields

$$|\Delta_A\rangle = \sum_{B:B \neq A} \frac{\lambda_B}{\lambda_A - \lambda_B} \epsilon_{BA} X |\phi_B\rangle + O(D^{3/2}/d). \quad (5.20)$$

In particular, the leading term in (5.19) is sufficient to justify the above assumption stated below (5.17). This establishes the error bound stated in (5.17) and thus also in (5.20).

To leading order in the small parameters ϵ_{AB} and ϵ_A we then obtain

$$\langle \tilde{\Phi}_A | X | \phi_B \rangle_p = ({}_p\langle \phi_A | X^\dagger + {}_p\langle \Delta_A |) X | \phi_B \rangle = \delta_{AB} + \epsilon_{AB} + \frac{\lambda_B}{\lambda_A - \lambda_B} \epsilon_{AB} (1 - \delta_{AB}) + O(D/d). \quad (5.21)$$

Similarly, for any fixed state $|\phi_0\rangle$, we have

$$\langle \Delta_A | X P_0 | \phi_0 \rangle_p = O(D^{3/2}/d). \quad (5.22)$$

We can also use the above results to extract leading order corrections for the projection $\tilde{P}_0$ onto the kernel of $\tilde{H}$. In particular, we find

$$\begin{aligned} \tilde{P}_0 &:= P_{\mathcal{Y}} - \sum_A \frac{X |\Phi_A\rangle {}_p\langle \Phi_A | X^\dagger}{\langle \tilde{\Phi}_A | \tilde{\Phi}_A \rangle_{\text{np}}} \\ &= P_{\mathcal{Y}} - \sum_A X \left[(1 - \epsilon_{AA}) | \phi_A \rangle {}_p\langle \phi_A | + \sum_{B:B \neq A} \frac{\lambda_B}{\lambda_A - \lambda_B} (\epsilon_{BA} | \phi_B \rangle {}_p\langle \phi_A | + \epsilon_{AB} | \phi_A \rangle {}_p\langle \phi_B |) \right] X^\dagger \\ &\quad + O\left(\frac{D^{3/2}}{d}\right). \end{aligned} \quad (5.23)$$

In contrast, the projection onto the kernel of H is $P_0 = \mathbb{1}_p - \sum_A | \phi_A \rangle {}_p\langle \phi_A |$.

5.2 Evolutions of a generic state

With the above results in hand, we can now compare the actions of $E_1(t)$, $E_2(t)$ on a generic state $|\phi\rangle$. We remind the reader that we use the term ‘generic’ to refer to states that are selected independently of the random variables in the Wishart distribution. Since we assumed the form of (5.5) to be fixed (in the sense that it is independent of d), we have effectively required D to be of order one as $d \rightarrow \infty$. We will thus cease to keep track of factors of D in the error terms below.

Let us write our state in the form

$$|\phi\rangle = P_0 |\phi\rangle + \sum_{A=1}^D c_A |\phi_A\rangle. \quad (5.24)$$

Perturbative evolution followed by acting with X thus yields

$$E_1(t) |\phi\rangle = X e^{-iHt} |\phi\rangle = X P_0 |\phi\rangle + \sum_A c_A e^{-i\lambda_A t} X |\phi_A\rangle. \quad (5.25)$$

On the other hand, acting with X and then evolving non-perturbatively gives

$$\begin{aligned}
 E_2(t)|\phi\rangle &= e^{-i\tilde{H}t}X|\phi\rangle = \tilde{P}_0X|\phi\rangle + \sum_A e^{-i\tilde{\lambda}_At} \frac{X|\Phi_A\rangle_{\text{p}}\langle\Phi_A|X^\dagger}{\langle\tilde{\Phi}_A|\tilde{\Phi}_A\rangle_{\text{np}}}X|\phi\rangle \\
 &= \tilde{P}_0X|\phi\rangle + \sum_A e^{-i\tilde{\lambda}_At} \frac{X|\Phi_A\rangle_{\text{p}}\langle\Phi_A|X^\dagger}{\langle\tilde{\Phi}_A|\tilde{\Phi}_A\rangle_{\text{np}}}XP_0|\phi\rangle + \sum_{AB} c_B e^{-i\tilde{\lambda}_At} \frac{X|\Phi_A\rangle_{\text{p}}\langle\Phi_A|X^\dagger}{\langle\tilde{\Phi}_A|\tilde{\Phi}_A\rangle_{\text{np}}}X|\phi_B\rangle \\
 &= XP_0|\phi\rangle + \sum_A c_A X|\phi_A\rangle - \sum_A \frac{X|\Phi_A\rangle_{\text{p}}\langle\Phi_A|X^\dagger}{\langle\tilde{\Phi}_A|\tilde{\Phi}_A\rangle_{\text{np}}}X \left(P_0|\phi\rangle + \sum_B c_B |\phi_B\rangle \right) \\
 &\quad + \sum_A \frac{X|\Phi_A\rangle_{\text{p}}\langle\Phi_A|X^\dagger}{\langle\tilde{\Phi}_A|\tilde{\Phi}_A\rangle_{\text{np}}} e^{-i\tilde{\lambda}_At} \left(XP_0|\phi\rangle + \sum_B c_B X|\phi_B\rangle \right).
 \end{aligned} \tag{5.26}$$

We can now use the results of section 5.1 to further evaluate the final line of (5.26). This yields

$$\begin{aligned}
 &\sum_A e^{-i\tilde{\lambda}_At} \left(\varepsilon_A X|\phi_A\rangle + c_A(1 - \epsilon_{AA})X|\phi_A\rangle + \sum_B \epsilon_{AB}c_B X|\phi_A\rangle \right. \\
 &\quad \left. + \sum_{B:B \neq A} \frac{\lambda_B}{\lambda_A - \lambda_B} \left[c_A \epsilon_{BA} X|\phi_B\rangle + c_B \epsilon_{AB} X|\phi_A\rangle \right] \right) + O\left(\frac{1}{d}\right) \\
 &= \sum_A \left(e^{-i\tilde{\lambda}_At} (c_A + \varepsilon_A) + \sum_{B:B \neq A} c_B \frac{\lambda_A}{\lambda_A - \lambda_B} \epsilon_{AB} (e^{-i\tilde{\lambda}_At} - e^{-i\tilde{\lambda}_Bt}) \right) X|\phi_A\rangle + O\left(\frac{1}{d}\right).
 \end{aligned} \tag{5.27}$$

We can then similarly expand the other terms in (5.26). For example, using (5.23) we find

$$\begin{aligned}
 &\frac{X|\Phi_A\rangle_{\text{p}}\langle\Phi_A|X^\dagger}{\langle\tilde{\Phi}_A|\tilde{\Phi}_A\rangle_{\text{np}}} \\
 &= X \left[(1 - \epsilon_{AA}) |\phi_A\rangle_{\text{p}}\langle\phi_A| + \sum_{B:B \neq A} \frac{\lambda_B}{\lambda_A - \lambda_B} (\epsilon_{BA} |\phi_B\rangle_{\text{p}}\langle\phi_A| + \epsilon_{AB} |\phi_A\rangle_{\text{p}}\langle\phi_B|) \right] X^\dagger + O\left(\frac{1}{d}\right).
 \end{aligned} \tag{5.28}$$

From this, it follows that

$$\frac{X|\Phi_A\rangle_{\text{p}}\langle\Phi_A|X^\dagger}{\langle\tilde{\Phi}_A|\tilde{\Phi}_A\rangle_{\text{np}}}XP_0|\phi\rangle = \varepsilon_A X|\phi_A\rangle + O\left(\frac{1}{d}\right) \tag{5.29a}$$

and

$$\begin{aligned}
 &\frac{X|\Phi_A\rangle_{\text{p}}\langle\Phi_A|X^\dagger}{\langle\tilde{\Phi}_A|\tilde{\Phi}_A\rangle_{\text{np}}}X|\phi_C\rangle = \delta_{AC}X|\phi_A\rangle + (1 - \delta_{AC})\epsilon_{AC} \frac{\lambda_A}{\lambda_A - \lambda_C} X|\phi_A\rangle \\
 &\quad + \sum_{B:B \neq A} \frac{\lambda_B}{\lambda_A - \lambda_B} \epsilon_{BA} \delta_{AC} X|\phi_B\rangle + O\left(\frac{1}{d}\right).
 \end{aligned} \tag{5.29b}$$

Substituting these results into (5.26) we obtain

$$(E_2(t) - E_1(t))|\phi\rangle = \sum_{A=1}^D \delta_A X|\phi_A\rangle, \tag{5.30}$$

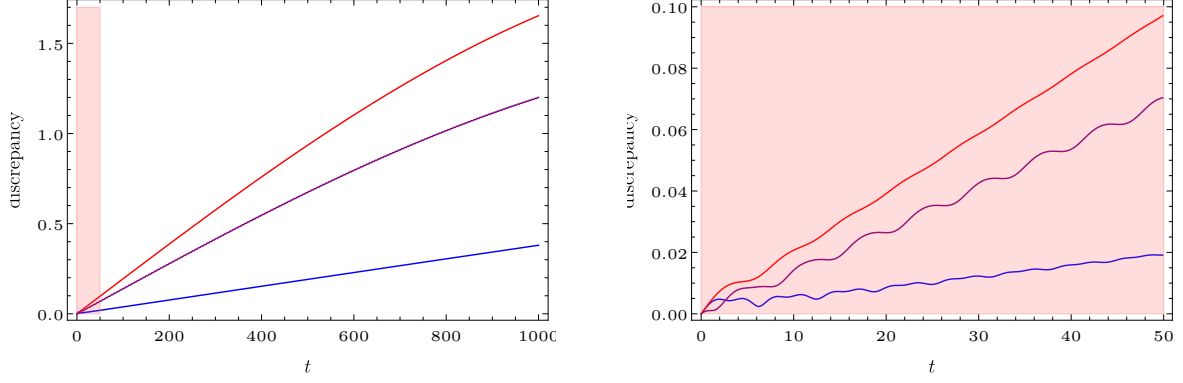


Figure 7. The discrepancy (5.35) for a small-rank Hamiltonian ($k = 2d, d = 10^6, \sqrt{d} = 1000, D = 4$) with eigenvalues $\lambda_1 = 0.5, \lambda_2 = 1.5, \lambda_3 = 2.5, \lambda_4 = 3.5$ and a random non-perturbative inner product drawn from the Wishart distribution (2.16). Results are shown for the second perturbative eigenstate $|\phi_2\rangle$ (lower curve in blue), the third perturbative eigenstate $|\phi_3\rangle$ (upper curve in red), and the linear combination $\frac{1}{\sqrt{2}}(|\phi_2\rangle + |\phi_3\rangle)$ of these eigenstates (middle curve in purple). As emphasized in the right panel, the discrepancy is small for $t \ll \sqrt{d}$, but it becomes of order 1 when $t \sim \sqrt{d}$ as shown on the left. There are $O(1)$ variations between random draws from the Wishart ensemble, though the qualitative form shown is common to many such draws.

where

$$\delta_A = c_A \left(e^{-i\tilde{\lambda}_A t} - e^{-i\lambda_A t} \right) + \left(e^{-i\tilde{\lambda}_A t} - 1 \right) \epsilon_A + \sum_{B: B \neq A} c_B \frac{\lambda_A}{\lambda_A - \lambda_B} \epsilon_{AB} (e^{-i\tilde{\lambda}_A t} - e^{-i\tilde{\lambda}_B t}). \quad (5.31)$$

From (5.17) we have $\tilde{\lambda}_A - \lambda_A = \mathcal{O}(1/\sqrt{d})$, so that the first term in (5.31) may be rewritten in the form

$$c_A e^{-i\tilde{\lambda}_A t} \left(1 - e^{i(\tilde{\lambda}_A - \lambda_A)t} \right) = c_A O\left(t[\lambda_A - \tilde{\lambda}_A]\right) = c_A O\left(\frac{t}{\sqrt{d}}\right). \quad (5.32)$$

Thus for $t \ll \sqrt{d}$ this term remains very small.

The second term in (5.31) is a bounded function of time with amplitude of order $\frac{1}{\sqrt{d}}$. However, the third term is more complicated. It clearly becomes large for $\lambda_A - \lambda_B$ small, so let us assume these differences to be $O(1)$. The term then involves a sum over a set of bounded functions of time with coefficients of order $c_B/\sqrt{d}$. The functions of time are small when t is small, after which they will generically contribute random phases. Thus we see that this sum will again be of order $\frac{\sqrt{\sum_B |c_B|^2}}{\sqrt{d}} < 1/\sqrt{d}$ so that we find

$$\delta_A = O(1/\sqrt{d}) + c_A O(t/\sqrt{d}) \quad \text{for } t < O(\sqrt{d}). \quad (5.33)$$

As a result, for any generic state $|\phi\rangle$ we find

$$(E_2(t) - E_1(t))|\phi\rangle = O(\sqrt{1/d}) + O(t/\sqrt{d}) \quad \text{for } t < O(\sqrt{d}). \quad (5.34)$$

A numerical example is shown in figure 7. For this purpose we define a real-valued measure of the discrepancy between $E_1(t)$ and $E_2(t)$ acting on the given state $|\phi\rangle$:

$$\text{discrepancy} \equiv \frac{\|(E_1(t) - E_2(t))|\phi\rangle\|}{\|X|\phi\rangle\|}. \quad (5.35)$$

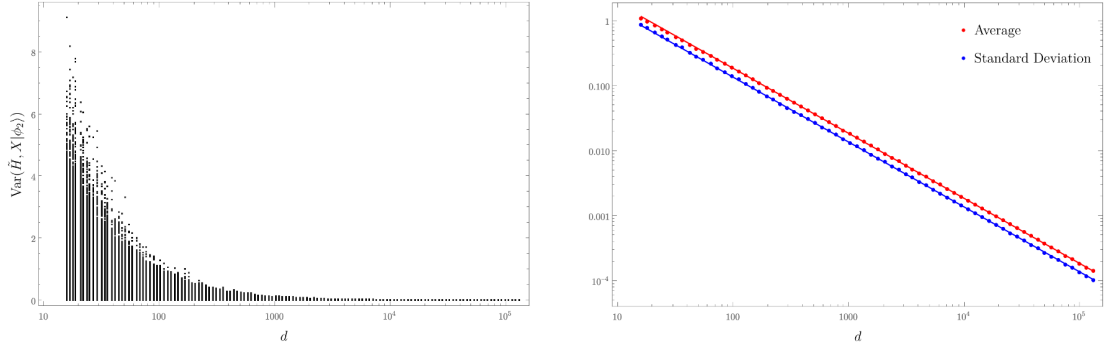


Figure 8. For each d , 10 000 random draws from the Wishart ensemble for $k = 2d$ were used to compute the quantity $\text{var}(\tilde{H}, X|\phi_2\rangle) := \frac{\langle \phi_2 | X^\dagger X H X X^\dagger X | \phi_2 \rangle}{\langle \phi_2 | X^\dagger X | \phi_2 \rangle} - \left(\frac{\langle \phi_2 | X^\dagger X H X | \phi_2 \rangle}{\langle \phi_2 | X^\dagger X | \phi_2 \rangle} \right)^2$ for the $D = 4$ Hamiltonian with $\lambda_1 = 0.5, \lambda_2 = 1.5, \lambda_3 = 2.5, \lambda_4 = 3.5$. For each draw, this quantity encodes the variance (squared uncertainty) of XHX in the state $X|\phi_2\rangle$. Since $X|\phi_2\rangle$ becomes an eigenstate at large d , $\text{var}(H, X|\phi_2\rangle)$ should tend to 0 with probability one as $d \rightarrow \infty$ (with an error of order $1/d$). In the left panel, the data is shown in black dots, which indeed asymptotes to 0 as $d \rightarrow \infty$. In the right panel, the red and blue dots show the average value and standard deviation of the data in the left panel for each d . The red and blue solid lines are the best fit of the data for $d \geq 512$ to a curve of the form $A \cdot d^{-B}$. We find $B \sim 1.000$ and $B \sim 0.998$ for the average and standard deviation, respectively.

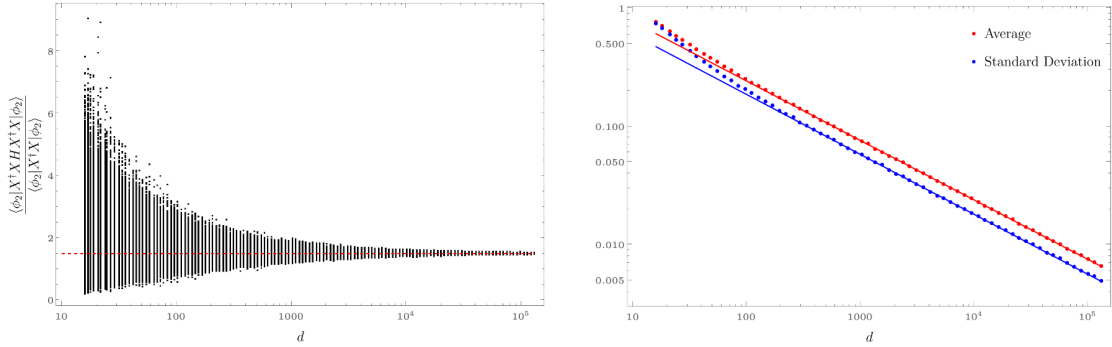


Figure 9. For each value of d shown, the 10,000 random draws from the Wishart ensemble from figure 8 are used to compute $\frac{\langle \phi_2 | X^\dagger X H X X^\dagger X | \phi_2 \rangle}{\langle \phi_2 | X^\dagger X | \phi_2 \rangle}$ for the same H as in figure 8. In the left panel, the data is shown in black dots and the dotted red line indicates the eigenvalue $\lambda_2 = 1.5$ of H to which the black dots should converge at large d . In the right panel, the red and blue dots show the average value and standard deviation of the absolute value of the residuals from the left panel for each d . The red and blue solid lines are the best fit of the data for $d \geq 512$ to a curve of the form $A \cdot d^{-B}$. We find $B \sim 0.503$ for the average and $B \sim 0.507$ for the standard deviation,.

We also include figures 8 and 9 which show numerical studies of the convergence of the eigenvectors of $XHX^\dagger$ to $X|\phi_A\rangle$ and the associated eigenvalues to λ_A . The results support the claims that $|\Delta_A\rangle$ and $\delta\lambda$ are of order $1/\sqrt{d}$.

In fact, the discrepancy is even smaller than that described above when $|\phi\rangle$ is a randomly chosen state in $\mathcal{H}_p$ as then $\sum_{A=1}^D |c_A|^2 = O\left(\frac{D}{k}\right)$. However, in that case the improvement is due to the fact that, because H was chosen to be of small rank, most of the state simply does not evolve at all under either E_1 or E_2 . Such states may thus be of less interest.

As described in section 3.3, our formalism also allows us to discuss perturbative Hamiltonians H that couple our topological gravity model to a non-gravitational ‘bath’ system with Hilbert space $\mathcal{H}_{\text{ext}}$. We can apply the above analysis directly to this case when there is a D -dimensional subspace $\mathcal{H}_{\text{dynamical}} \subset \mathcal{H}_{\text{p}}$ such that H acts nontrivially only on $\mathcal{H}_{\text{dynamical}} \otimes \mathcal{H}_{\text{ext}} \subset \mathcal{H}_{\text{np}} \otimes \mathcal{H}_{\text{ext}}$. For $D \ll d$ the perturbative and non-perturbative evolutions will again agree for $t \ll \sqrt{d}$. But when the smallest such D is instead comparable to d , the evolutions will generally show marked differences even for short times — at least in randomly-chosen initial states.⁸

5.3 Degeneracies and approximate degeneracies

While the denominators in (5.31) diverge when two eigenvalues of H coincide, such exact degeneracies are not in fact hard to handle. As usual in degenerate perturbation theory, given an n -fold degenerate eigenspace of H we may simply choose the orthonormal basis $|\phi_1\rangle, |\phi_2\rangle, \dots, |\phi_n\rangle$ in such a way that each associated $|\Delta_A\rangle$ ($A \in \{1, \dots, n\}$) is orthogonal to all $X|\phi_B\rangle$ with $B \in \{1, \dots, n\}$ and not just for $B = A$. This choice removes all divergent denominators from (5.31) so that the analysis of section 5.2 continues to apply.

The case where eigenvalue differences $\lambda_A - \lambda_B$ within each band become small as $d \rightarrow \infty$ is difficult to analyze in general. This case also seems rather artificial as the perturbative Hamiltonian by definition knows nothing about the non-perturbative corrections that set the dimension of $\mathcal{H}_{\text{np}}$ to d .

However, we mention briefly that we can in fact address the case where H has bands of eigenvalues within each of which the eigenvalue differences $\lambda_A - \lambda_B$ become small as $d \rightarrow \infty$ but where distinct bands remain separated by $O(1)$ gaps. In this case we can write H as the sum of an exactly-degenerate Hamiltonian H_0 (whose eigenvalues label the above bands) and a term V that lifts the degeneracies but leaves the splittings small. If we assume the small splittings to be of order $1/\sqrt{d}$, then V is also of order $1/\sqrt{d}$. As a result, up to times $t \ll \sqrt{d}$, H_0 gives a good approximation to the evolution given by H . Furthermore, the corresponding non-perturbative Hamiltonians $\tilde{H}_0 = XH_0X^\dagger$ and $\tilde{H} = XHX^\dagger$ will also generate nearly-identical time evolutions for times $t \ll \sqrt{d}$. And since we have just seen that the analysis of section 5.2 can be applied to H_0 and $\tilde{H}_0$, it follows that the perturbative and non-perturbative evolutions E_1, E_2 defined by H also agree well for times $t = O(d^\alpha)$ for $\alpha < \frac{1}{2}$.

5.4 Finely-tuned states with large discrepancies

An important caveat in the above analysis is that the high probability of long-time agreement between $E_1(t)$ and $E_2(t)$ derived in section 5.2 holds only for what we called *generic* states. These were defined to be states chosen independently of the $\tilde{\alpha}$ -sector drawn from the Wishart distribution; i.e., the state was chosen without knowledge of the random variable X . However, as we now show, there is also a D -dimensional X -dependent subspace of $\mathcal{H}_{\text{p}}$ on which the two operators differ significantly. This subspace is just the image of M or, equivalently, the image of X . Indeed, let us consider the space spanned by the $X|\phi_A\rangle$, where $|\phi_A\rangle$ again represent

⁸In cases where there is a subspace $\mathcal{H}_{\text{small}} \subset \mathcal{H}_{\text{ext}}$ such that H preserves the space $\mathcal{H}_{\text{eff}} := \mathcal{H}_{\text{p}} \otimes \mathcal{H}_{\text{small}}$, and where the restriction of H to $\mathcal{H}_{\text{eff}}$ gives a Hamiltonian with a smaller value of D , the discrepancy can be reduced by choosing the initial state to lie in $\mathcal{H}_{\text{eff}}$.

the eigenvectors of H with non-vanishing eigenvalues. It was already argued in section 5.1 that $X|\phi_A\rangle$ is an approximate eigenvector of $\tilde{H}$ with eigenvalue $\tilde{\lambda}_A \approx \lambda_A$. Furthermore, in the limit $\frac{k}{d} \rightarrow \infty$, appendix A.3 shows that X satisfies $X^\dagger X = X^2 = \sqrt{\frac{k}{d}}X$, which then yields

$$E_2(t)X|\phi_A\rangle \approx \sqrt{\frac{k}{d}}e^{-i\lambda_A t}X|\phi_A\rangle. \quad (5.36)$$

In contrast, let us consider the perturbative evolution e^{-iHt} . Since X is a uniformly-chosen random projection, $X|\phi_A\rangle$ will have only a tiny correlation with $|\phi_A\rangle$ in the limit $k \gg d$. In particular, in that limit $X|\phi_A\rangle$ will have amplitude 1 to be orthogonal to $|\phi_A\rangle$. And since $k \gg d \gg D$ it will also have amplitude 1 to lie in $\ker H$. Thus,

$$E_1(t)X|\phi_A\rangle \approx \sqrt{\frac{k}{d}}X|\phi_A\rangle. \quad (5.37)$$

The two evolutions thus differ by a large phase. Of course, a global phase is of no physical relevance. However, different $X|\phi_A\rangle$ clearly lead to different phases, so physical differences between E_1 and E_2 become large for superpositions of these states.

We thus conclude that $E_2(t) - E_1(t)$ is not small on the subspace spanned by the $X|\phi_A\rangle$. A certain caution is thus needed in distinguishing results that hold with high probability for states chosen independently of X from those that hold for all possible states. We have included a numerical example in figure 10. We may also read this result backwards to say that, for any state $|\psi\rangle$, there are α -sectors in which $(E_1(t) - E_2(t))|\psi\rangle$ is large, namely those for which we have

$$|\psi\rangle = \frac{X|\psi_1\rangle}{\sqrt{\langle\psi_1|X^\dagger X|\psi_1\rangle}} \quad (5.38)$$

for some state $|\psi_1\rangle$.

6 Discussion

The above work investigated the effect of null states on time-evolutions defined by gravitational path integrals. We worked in a simple topological model of quantum gravity with k end-of-the-world branes, and we focused on the one-boundary sector of such models. Since the model is topological, the source-free Hamiltonian vanishes identically. However, general time-dependence was introduced by adding boundary sources that insert one-boundary operators defined by annihilating and creating end-of-the-world branes. Such insertions led to a non-vanishing Hamiltonian H at the perturbative level, as well as a non-perturbative Hamiltonian $\tilde{H}$. In this context, even for times that are short in comparison with natural timescales of the perturbative Hamiltonian H , we found the perturbative and non-perturbative time evolutions defined by such path integrals to have large corrections when the rank D of H is of the same order as the dimension d of the non-perturbative Hilbert space, and to be starkly different for $D \gg d$. On the other hand, for $D \ll d$ the perturbative and non-perturbative evolutions are nearly identical for all times t of order d^α with $\alpha < \frac{1}{2}$. We emphasize that such timescales are exponentially large in the entropy of the system. In particular, for $D \ll d$ we showed that

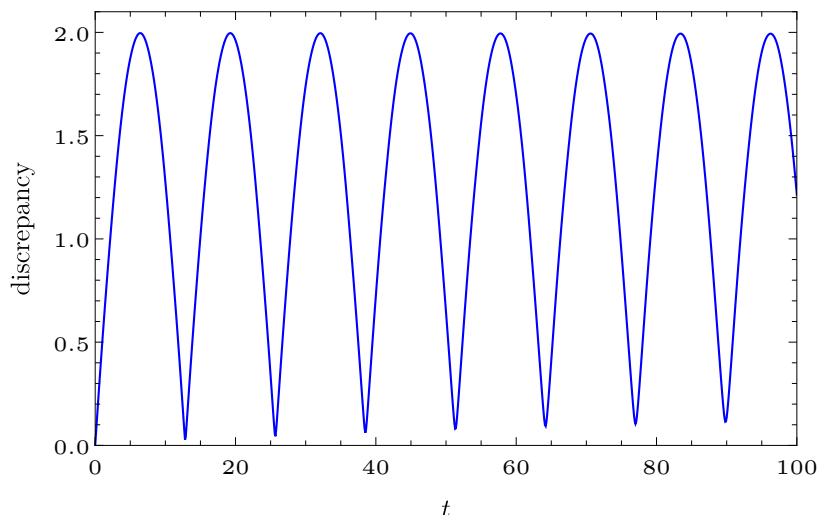


Figure 10. The figure shows the large discrepancy (5.35) that arises for a finely-tuned state $M|\phi_1\rangle$ with parameters $d = 10^3, k = 10^6, D = 4$. Here the eigenvalues of the perturbative Hamiltonian H are $\lambda_1 = 0.5, \lambda_2 = 1.5, \lambda_3 = 2.5, \lambda_4 = 3.5$. The spacing between the peaks agrees well with $2\pi/\lambda_1 = 4\pi$ as predicted by comparing (5.36) with (5.37).

the operation $E_1(t)$, defined by applying the perturbative time-evolution to a perturbative state and then mapping the result to the non-perturbative Hilbert space, was an excellent approximation to the operation $E_2(t)$ defined by first mapping the perturbative state to the non-perturbative Hilbert space and then evolving with non-perturbative time evolution. Our formalism also allowed us to discuss dynamics that couples our topological gravity system to a non-gravitational ‘bath’ system, in which case analogous results apply.

A critical point is that, even though $E_1(t)$ and $E_2(t)$ differ strongly for $D \gtrsim d$, the non-perturbative dynamics is manifestly unitary for any value of D . Furthermore, this remains true when the system is coupled to an external bath. On the other hand, section 7.3 of [9] argued in a related context that interactions with a system with null states was necessarily non-unitary. The scenario considered by [9] involved a large parent universe that emits a set of baby universes, and where the pieces to be emitted are in a state that — if emitted as baby universes — would become a null state. Thus the emission process appeared to take a state of finite-norm to one of zero norm, violating unitarity in a particularly strong way.

To resolve this tension, let us compare the discussion of [9] with our work above. Since the full time evolution — including any emissions of baby universes — must be defined by a non-perturbative path integral, even if one described the initial state in a perturbative language we would expect the resulting operation to correspond to something like our $E_2(t)$ which continually incorporates the effects of wormholes and α -sectors at all times t . In contrast, [9] mentions possible effects of null states only *after* the baby universes have been emitted from the parent universe. It thus assumes that the emission process itself can be described in terms of states in the perturbative Hilbert space. It is only when the process is completed that [9] evaluates the result using the non-perturbative inner product. It thus appears to us that the notion of time-evolution described in [9] corresponds to something like our $E_1(t)$, rather than to our $E_2(t)$. And while it is certainly true that $E_1(t)$ is generally far

from unitary, we have seen that there is no tension with the unitarity of $E_2(t)$. Indeed, our interpretation of the scenario described in [9] is that, since the non-perturbative evolution *must* remain unitary in the presence of null states, the non-perturbative time evolution is *required* to deviate strongly from the perturbative evolution. Strong deviations were indeed seen in section 4 for our model with $D \gg d$, though of course many details remain to be understood for more realistic contexts.

While our calculations were performed for time-independent Hamiltonians, the evolution operator $U(t) = \mathcal{P} \exp(-i \int H(t))$ for time-dependent $H(t)$ can be well-approximated by a product of evolutions over short time intervals, each of which is defined by a Hamiltonian that is time-independent within the given interval. It is thus clear that the perturbative and non-perturbative evolutions are again nearly identical over spans of time in which the non-trivial part of $H(t)$ is confined to a D -dimensional subspace of the perturbative Hilbert space $\mathcal{H}_p$ with $D \ll d$, though the discrepancy should again become large once $H(t)$ explores a subspace with $D \gtrsim d$.

Although the small D agreement is technically interesting, our most striking result may be the large discrepancy that arises between the perturbative and non-perturbative time evolutions for $D \gtrsim d$. Recall, for example, that even in asymptotically AdS space any small band of energies is associated with an infinite set of states in the perturbative gravity description. In particular, due to the possibility of having a black hole with a long throat deep inside the horizon, this is the case even if we require states to have no structure within a Planck distance of any event horizon (i.e., if we impose some sort of stretched horizon cutoff). One might thus attempt to model such settings by taking the $k \rightarrow \infty$ limit of our results. Furthermore, since all such states have similar non-zero energies E , it is clear that the perturbative Hamiltonian has full-rank on this space. But path integral computations (see e.g. [3, 4, 6–8, 17]) indicate that the non-perturbative theory in such contexts has only a finite number of states $d = e^{S_{\text{BH}}}$, where S_{BH} is the associated Bekenstein-Hawking entropy. Carrying over our results directly would then suggest that the non-perturbative dynamics in such contexts bears no relation whatsoever to the perturbative dynamics, and that it is instead essentially random.

Although our model showed no hints that quantum extremal surfaces are important for the above effect, one might nevertheless hope that in interesting gravitational theories any such effects may be confined to the interior of black holes (or, perhaps, to the regions inside and just outside an event horizon in the context described by [18]). Furthermore, the random dynamics seen in our model appears similar in spirit to the idea that there may be no coherent semiclassical spacetime inside a black hole associated with the firewall hypothesis of [19]. But our model suffers from an especially extreme form of such effects which appear to arise equally for ‘young’ black holes⁹ as well as those that are ‘old.’ The reader may thus naturally find such a scenario to be unpalatable.

A potential clue to how this predicament might be avoided can be found by recalling that we induced time-evolution in our model by inserting operators, and that we found the above results even for insertions that give $H = c\mathbb{1}_p$ where $c \in \mathbb{R}$ and $\mathbb{1}_p$ is the identity on the

⁹I.e., even for special states confined to tiny sectors of the Hilbert space which we again choose without knowledge of the random draw M from the Wishart ensemble.

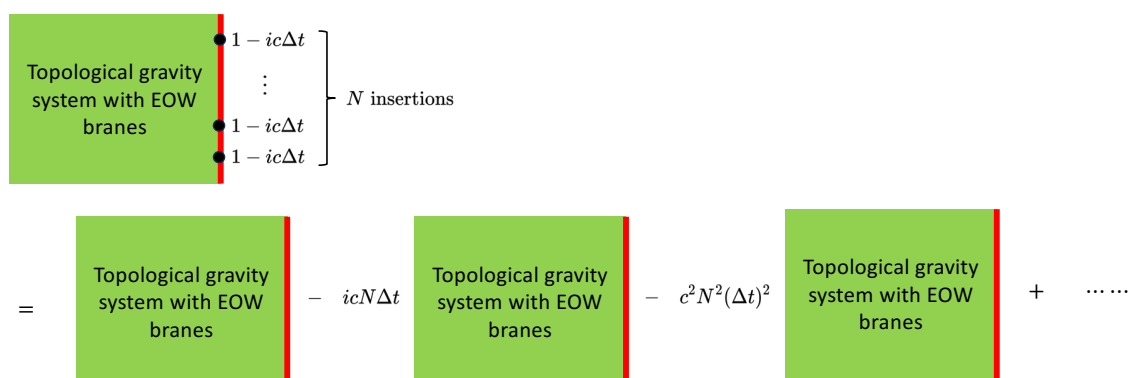


Figure 11. N insertions of the c-numbers $(1 - ic\Delta t)$ are equivalent to summing a series of path integrals with trivial insertions and non-trivial weights.

perturbative Hilbert space. However, for this special case there is an alternative way to insert the operator $iH\Delta t$, which is to simply add a term with a single asymptotic boundary weighted by $ic\Delta t$ as shown in the 2nd term on the right in figure 11. For past boundary condition $|I\rangle$ and future boundary condition $|J\rangle$, the perturbative evaluation of the one-insertion path integral then yields precisely $ict\langle I|\mathbb{1}_p|J\rangle_p = ict\langle I|J\rangle_p$ as desired. However, instead of the matrix elements of a random operator implied by the results found in section 5, it is immediately clear that the one-insertion non-perturbative path integral computes precisely $ict\langle \tilde{I}|\tilde{J}\rangle_{np}$. The corresponding non-perturbative Hamiltonian is thus $\tilde{H} = c\mathbb{1}_{np}$, which is now distinctly different from the operator $XHX^\dagger$.

This, then, suggests that it is important to discuss such issues directly in the language of boundary conditions rather than using the language of operators and states on the perturbative Hilbert space. We will return to this point in a forthcoming work. However, for the moment we simply observe that a given model may feature certain boundary conditions for which the perturbative and non-perturbative evolutions agree to good precision for long times and which we may call np-stable. In our topological model, the np-stable boundary conditions include the trivial boundary condition (representing the identity) and boundary conditions describing the annihilation and creation of small numbers of EOW-branes. A tantalizing fantasy is that for higher-dimensional AdS/CFT models the np-stable boundary conditions might include those with sufficiently smooth boundary-sources of low conformal dimension (i.e., sources that define relevant or marginal deformations). In this context, the long-time agreement we found for low-rank Hamiltonians may naturally generalize to similar successes for linear combinations of a small number of np-stable boundary conditions. However, unless the boundary condition defines a Hamiltonian that approximates a linear combination of the identity operator and an operator of small rank, the potentially-vast difference in dimension between the perturbative and non-perturbative Hilbert spaces makes it unlikely for the perturbative and non-perturbative evolutions to act similarly across the entire space of available states. In particular, the recurrence time for the non-perturbative evolution will generally be much shorter. One might thus expect that a useful notion of np-stability applies only when evolving a small and special space of states which might, for

example, be associated with appropriately ‘young’ black holes. We hope to return to this issue in future work as well by studying non-perturbative time evolution in more realistic models, perhaps based on recent analyses of non-perturbative geodesic lengths in JT gravity [20].

As a final comment, we note that the reader may find it tempting to associate the special np-stable boundary conditions with the notion of ‘simple operators’ described by Akers, Engelhardt, Harlow, Penington, and Vardhan in their discussion [21] of non-isometric codes. While it would certainly be interesting to explore such connections, we emphasize that our proposed notion of np-stable boundary conditions is intended to mean that the perturbative and non-perturbative path integrals both define *linear* operators with compatible time evolutions (perhaps on a small linear space of states).

In contrast, the non-isometric code paradigm of [21] would in principle allow us to declare in our topological model that the identity operator remains simple when represented in the form $H = \sum_I |I\rangle_p \langle I|$, and to then recover the associated (trivial) perturbative dynamics from the action of the operator $XHX^\dagger$ discussed in section 5. A similar phenomenon occurs in the dynamical model of [21] where the non-perturbative evolution is simply reinterpreted as a non-isometric map applied to the desired perturbative evolution. While the non-isometric codes of [21] are generally constrained by the extrapolate dictionary, such constraints become trivial in models like the one studied above that lack propagating bulk fields. And even in more complete models, this dictionary imposes no obvious constraints inside the horizon. The freedom to choose a general non-isometric map V thus allows one to essentially define V to be equivariant with respect to the perturbative and non-perturbative time evolutions even if the two evolutions have no common structure deep in the bulk. The resulting flexibility makes the non-isometric code paradigm extremely powerful, though at the apparent cost of making the perturbative evolution (called the effective description in [21]) essentially independent of the fundamental non-perturbative evolution. We would prefer to instead see any effective description derived directly from a fundamental description, rather than have the two simply rendered consistent by fiat. Again, we hope to return to such issues in future work.

Acknowledgments

The authors thank Geoff Penington for discussions of non-isometric codes. The work of XD was supported in part by the U.S. Department of Energy, Office of Science, Office of High Energy Physics, under Award Number DE-SC0011702, and by funds from the University of California. The work of MK, XL, DM, and ZW was supported by NSF grant PHY-2107939 and by funds from the University of California. ZW was also supported by Air Force Office of Scientific Research under award number FA9550-19-1-036 and by the DOE award number DE-SC0015655. This research was supported in part by grant NSF PHY-2309135 to the Kavli Institute for Theoretical Physics (KITP).

A Wishart distribution

This appendix summarizes our conventions regarding the Wishart distribution and performs a few calculations. Other calculations described in the main text are similar.

A.1 Basic definitions

The Wishart distribution is a unitarily invariant distribution on the space of $k \times k$ hermitian linear matrices. Let us introduce a set of gaussian variables a_{Ii} satisfying

$$\langle a_{Ii} \rangle = 0, \quad \langle a_{Ii} a_{Jj} \rangle = 0, \quad \langle a_{Ii}^* a_{Jj}^* \rangle = 0, \quad \langle a_{Ii}^* a_{Jj} \rangle = \delta_{IJ} \delta_{ij}, \quad (\text{A.1})$$

where $I, J \in \{1, \dots, k\}$ and $i, j \in \{1, \dots, d\}$. All higher moments of the a_{Ii} may be calculated using Wick's theorem.

From the a_{Ii} , we may build a rectangular random matrix that maps from $\mathbb{C}^d$ to $\mathbb{C}^k$:

$$[G]_{Ii} = a_{Ii}. \quad (\text{A.2})$$

We may then construct a random $k \times k$ matrix M which is said to be drawn from the Wishart distribution:

$$M := \frac{1}{d} G G^\dagger. \quad (\text{A.3})$$

More explicitly, this matrix may be written

$$[M]_{IJ} = \frac{1}{d} \sum_{i=1}^d a_{Ii} a_{Ji}^*. \quad (\text{A.4})$$

The matrix M is hermitian by construction. Moreover, because it can be expressed as the product (A.3), we have

$$\text{rank } M \leq \min(d, k). \quad (\text{A.5})$$

The inequality is saturated with probability one as obtaining a smaller rank requires an accidental degeneracy.

A.2 A few moments

We will now explicitly calculate the first few moments of M . This is a nice warm-up before performing some of the more intricate calculations described in the main text.

A.2.1 Mean

We want to calculate $\langle M \rangle$. This is quite immediate, since we have:

$$[\langle M \rangle]_{IJ} = \frac{1}{d} \sum_{i=1}^d \langle a_{Ii} a_{Ji}^* \rangle = \frac{1}{d} \sum_{i=1}^d \delta_{IJ} \delta_{ii} = [\mathbb{1}_p]_{IJ}, \quad (\text{A.6})$$

and thus $\langle M \rangle = \mathbb{1}_p$.

A.2.2 Variance

To calculate the variance of M we will also need $\langle M^2 \rangle$. More explicitly, we wish to compute the ensemble average of the matrix

$$[M^2]_{IK} = \sum_{J=1}^k M_{IJ} M_{JK} = \frac{1}{d^2} \sum_{i,j=1}^d \sum_{J=1}^k \left(a_{Ii} a_{Ji}^* a_{Jj} a_{Kj}^* \right). \quad (\text{A.7})$$

Taking $\langle \cdot \rangle$ on both sides and applying Wick's theorem yields

$$\langle M^2 \rangle_{IK} = \frac{1}{d^2} \sum_{i,j=1}^d \sum_{J=1}^k \langle a_{Ii} a_{Ji}^* a_{Jj} a_{Kj}^* \rangle = \frac{1}{d^2} \sum_{i,j=1}^d \sum_{J=1}^k \left(\langle a_{Ii} a_{Ji}^* \rangle \langle a_{Jj} a_{Kj}^* \rangle + \langle a_{Ii} a_{Kj}^* \rangle \langle a_{Jj} a_{Ji}^* \rangle \right), \quad (\text{A.8})$$

and thus

$$\langle M^2 \rangle_{IK} = \frac{1}{d^2} \sum_{i,j=1}^d \sum_{J=1}^k (\delta_{IJ} \delta_{ii} \delta_{JK} \delta_{jj} + \delta_{IK} \delta_{ij} \delta_{JJ}) = \delta_{IK} + \frac{k}{d} \delta_{IK} = \left(1 + \frac{k}{d} \right) [\mathbb{1}_p]_{IK}. \quad (\text{A.9})$$

The variance is then $\langle M^2 \rangle - \langle M \rangle^2 = \frac{k}{d} \mathbb{1}_p$. We see that for $k \gg d$, the matrix M is far from being sharply peaked. This should be no surprise since, although the ensemble is invariant under $k \times k$ unitaries, any given M has rank $d \ll k$ and must thus determine a preferred d -dimensional subspace of a rank k vector space.

Another useful result is the two-point function $\langle M_{I_1 I_2} M_{I_3 I_4} \rangle$. Wick's contractions analogous to those above yield

$$\langle M_{I_1 I_2} M_{I_3 I_4} \rangle = [\mathbb{1}_p]_{I_1 I_2} [\mathbb{1}_p]_{I_3 I_4} + \frac{1}{d} [\mathbb{1}_p]_{I_1 I_4} [\mathbb{1}_p]_{I_2 I_3}. \quad (\text{A.10})$$

From (A.10) we can also see that for any (fixed) $k \times k$ matrix H we have:

$$\langle M H M \rangle = H + \frac{\text{Tr}_p H}{d} \mathbb{1}_p. \quad (\text{A.11})$$

In particular, taking $H = \mathbb{1}_p$ recovers our previous result.

A.3 Random projections

For large k , the central limit theorem tells us that the magnitudes of the vectors $\sum_I a_{Ii} |I\rangle$ should have small fluctuations. Furthermore, for $k \gg d$ we expect these d vectors to be approximately orthonormal. In this limit we thus expect M to be proportional to a rank d projection. If this is so, then the result (3.5) tells us that the projection must be $\mathcal{P} = \frac{d}{k} M$. Since the distribution for M is clearly invariant under unitary transformations, the range of $\mathcal{P}$ must be a random subspace with respect to the uniform distribution.¹⁰

We now explore the details of this argument. Since $\mathcal{P}$ is manifestly Hermitian, it is a projection if $\mathcal{P}^2 = \mathcal{P}$. A straightforward yet tedious calculations yields

$$\langle \mathcal{P}^2 - \mathcal{P} \rangle = \frac{d^2}{k^2} [\mathbb{1}_p], \quad (\text{A.12a})$$

¹⁰Geometrically, the space of all d -dimensional subspaces is $U(k)/(U(d) \times U(k-d))$. The uniform distribution on this space is given by the associated Haar measure.

while

$$\begin{aligned}\langle(\mathcal{P}^2 - \mathcal{P})^2\rangle &= \langle\mathcal{P}^4\rangle + \langle\mathcal{P}^2\rangle - 2\langle\mathcal{P}^3\rangle \\ &= \left(\frac{4d^2}{k^4} + \frac{d^4}{k^4} + \frac{3d}{k^3} + \frac{4d^3}{k^3} + \frac{2d^2}{k^2}\right)[\mathbb{1}_P].\end{aligned}\tag{A.12b}$$

In the limit $k/d \rightarrow \infty$ we thus find that $\mathcal{P}$ approaches a projection in the sense of the weak operator topology. However, a stronger measure of convergence is given by the Frobenius-norm, or the Schatten 1-norm:

$$\|Y\|^2 := \text{tr}(YY^\dagger),\tag{A.13}$$

which is easily computed by taking the trace of (A.12b). We find

$$\|\mathcal{P}^2 - \mathcal{P}\|^2 \sim \frac{2d^2}{k}.\tag{A.14}$$

So even in this stronger sense we find $\mathcal{P}$ to approach a projection in the stronger limit $k/d^2 \rightarrow \infty$.

Open Access. This article is distributed under the terms of the Creative Commons Attribution License ([CC-BY4.0](https://creativecommons.org/licenses/by/4.0/)), which permits any use, distribution and reproduction in any medium, provided the original author(s) and source are credited.

References

- [1] G. Penington, *Entanglement Wedge Reconstruction and the Information Paradox*, *JHEP* **09** (2020) 002 [[arXiv:1905.08255](https://arxiv.org/abs/1905.08255)] [[INSPIRE](#)].
- [2] A. Almheiri, N. Engelhardt, D. Marolf and H. Maxfield, *The entropy of bulk quantum fields and the entanglement wedge of an evaporating black hole*, *JHEP* **12** (2019) 063 [[arXiv:1905.08762](https://arxiv.org/abs/1905.08762)] [[INSPIRE](#)].
- [3] G. Penington, S.H. Shenker, D. Stanford and Z. Yang, *Replica wormholes and the black hole interior*, *JHEP* **03** (2022) 205 [[arXiv:1911.11977](https://arxiv.org/abs/1911.11977)] [[INSPIRE](#)].
- [4] A. Almheiri et al., *Replica Wormholes and the Entropy of Hawking Radiation*, *JHEP* **05** (2020) 013 [[arXiv:1911.12333](https://arxiv.org/abs/1911.12333)] [[INSPIRE](#)].
- [5] N. Engelhardt and A.C. Wall, *Quantum Extremal Surfaces: Holographic Entanglement Entropy beyond the Classical Regime*, *JHEP* **01** (2015) 073 [[arXiv:1408.3203](https://arxiv.org/abs/1408.3203)] [[INSPIRE](#)].
- [6] D. Marolf and H. Maxfield, *Transcending the ensemble: baby universes, spacetime wormholes, and the order and disorder of black hole information*, *JHEP* **08** (2020) 044 [[arXiv:2002.08950](https://arxiv.org/abs/2002.08950)] [[INSPIRE](#)].
- [7] V. Balasubramanian, A. Lawrence, J.M. Magan and M. Sasieta, *Microscopic Origin of the Entropy of Astrophysical Black Holes*, *Phys. Rev. Lett.* **132** (2024) 141501 [[arXiv:2212.08623](https://arxiv.org/abs/2212.08623)] [[INSPIRE](#)].
- [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** (2024) 011024 [[arXiv:2212.02447](https://arxiv.org/abs/2212.02447)] [[INSPIRE](#)].
- [9] B. Guo, M.R.R. Hughes, S.D. Mathur and M. Mehta, *Contrasting the fuzzball and wormhole paradigms for black holes*, *Turk. J. Phys.* **45** (2021) 281 [[arXiv:2111.05295](https://arxiv.org/abs/2111.05295)] [[INSPIRE](#)].

- [10] J.M. Maldacena and L. Maoz, *Wormholes in AdS*, *JHEP* **02** (2004) 053 [[hep-th/0401024](#)] [[INSPIRE](#)].
- [11] P. Saad, S.H. Shenker and D. Stanford, *JT gravity as a matrix integral*, [arXiv:1903.11115](#) [[INSPIRE](#)].
- [12] S.R. Coleman, *Black holes as red herrings: Topological fluctuations and the loss of quantum coherence*, *Nucl. Phys. B* **307** (1988) 867 [[INSPIRE](#)].
- [13] S.B. Giddings and A. Strominger, *Loss of incoherence and determination of coupling constants in quantum gravity*, *Nucl. Phys. B* **307** (1988) 854 [[INSPIRE](#)].
- [14] M. Reed and B. Simon, *Methods of Modern Mathematical Physics. Vol. I Functional Analysis*, Academic Press (1972).
- [15] E. Colafranceschi, X. Dong, D. Marolf and Z. Wang, *Algebras and Hilbert spaces from gravitational path integrals: Understanding Ryu-Takayanagi/HRT as entropy without invoking holography*, [arXiv:2310.02189](#) [[INSPIRE](#)].
- [16] G.T. Horowitz and J.M. Maldacena, *The black hole final state*, *JHEP* **02** (2004) 008 [[hep-th/0310281](#)] [[INSPIRE](#)].
- [17] D. Marolf and H. Maxfield, *Observations of Hawking radiation: the Page curve and baby universes*, *JHEP* **04** (2021) 272 [[arXiv:2010.06602](#)] [[INSPIRE](#)].
- [18] R. Bousso and G. Penington, *Islands Far Outside the Horizon*, [arXiv:2312.03078](#) [[INSPIRE](#)].
- [19] A. Almheiri, D. Marolf, J. Polchinski and J. Sully, *Black Holes: Complementarity or Firewalls?*, *JHEP* **02** (2013) 062 [[arXiv:1207.3123](#)] [[INSPIRE](#)].
- [20] L.V. Iliesiu et al., *On the non-perturbative bulk Hilbert space of JT gravity*, [arXiv:2403.08696](#) [[INSPIRE](#)].
- [21] C. Akers et al., *The black hole interior from non-isometric codes and complexity*, *JHEP* **06** (2024) 155 [[arXiv:2207.06536](#)] [[INSPIRE](#)].