Prepared for submission to JHEP

Transcending the ensemble: baby universes, spacetime wormholes, and the order and disorder of black hole information

Donald Marolf and Henry Maxfield

Department of Physics, University of California, Santa Barbara, CA 93106, USA

E-mail: marolf@physics.ucsb.edu, hmaxfield@physics.ucsb.edu

Abstract: In the 1980's, work by Coleman and by Giddings and Strominger linked the physics of spacetime wormholes to 'baby universes' and an ensemble of theories. We revisit such ideas, using features associated with a negative cosmological constant and asymptotically AdS boundaries to strengthen the results, introduce a change in perspective, and connect with recent replica wormhole discussions of the Page curve. A key new feature is an emphasis on the role of null states. We explore this structure in detail in simple topological models of the bulk that allow us to compute the full spectrum of associated boundary theories. The dimension of the asymptotically AdS Hilbert space turns out to become a random variable Z, whose value can be less than the naive number k of independent states in the theory. For $k > Z$, consistency arises from an exact degeneracy in the inner product defined by the gravitational path integral, so that many a priori independent states differ only by a null state. We argue that a similar property must hold in any consistent gravitational path integral. We also comment on other aspects of extrapolations to more complicated models, and on possible implications for the black hole information problem in the individual members of the above ensemble.

Contents

1 Introduction

The past year has seen several interesting developments in the study of black hole information. In particular, it has been well-known for some time that the von Neumann entropy $S_{\rm rad}$ of emitted Hawking radiation as a function of time gives an important diagnostic of whether and to what degree information is preserved or lost in evaporating black holes [1]. Familiar effective field theory would give an entropy that increases monotonically throughout the evaporation, even though the black hole's Bekenstein-Hawking entropy $S_{\rm BH} = \frac{A}{4G}$ monotonically decreases to a value near zero. In contrast, a model in which the black hole is a standard quantum system with density of states $S_{\rm BH}$ coupled unitarily to the radiation field would — when the initial state is pure — require $S_{\rm rad} \leq S_{\rm BH}$ at all times. As a result, in such models $S_{\rm rad}$ generally increases to a maximum, at which time it nearly equals $S_{\rm BH}$, and then decreases monotontically thereafter. The final phase with decreasing $S_{\rm rad}$ describes the return of information to the external universe from the black hole.

Despite many arguments suggesting that the latter so-called 'Page curve' should accurately approximate the result of black hole evaporation, for many years it was unclear how such a result could be obtained from a controlled gravitational calculation; see e.g. reviews in [2–6]. The plethora of proposals for new physics that might be associated with obtaining this Page curve (including [3, 7–39]) were thus all properly viewed as speculative and contained at least some optimistic extrapolation or ad hoc ingredient.[1]

Recently, however, it was noted that the 'unitary' Page curve, including the turnover of $S_{\rm rad}$, could be obtained by combining ideas from holography with effective field theory [42, 43] — or equivalently with quantum field theory in curved space. In particular, under very general conditions [42, 43] argued that one could obtain this result by computing the generalized entropy $S_{\rm gen} = \frac{A}{4G} + S_{\rm bulk}$ of an appropriate comdimension-2 quantum extremal surface (QES) , where the surface is chosen so that holography suggests this might represent $S_{\rm rad}$. Here $S_{\rm bulk}$ is the von Neumann entropy of bulk fields outside the codimension-2 QES. See also further explorations of this idea in [44–47].

Critically, [48, 49] then pointed out that — at least in some contexts — this seemingly-hybrid recipe in fact follows from replica trick calculations of $S_{\rm rad}$ using the gravitational path integral (and in particular that this was implicit in earlier derivations of the quantum corrected Ryu-Takayanagi [50, 51] and Hubeny-Rangamani-Takayanagi [52] entropy formulae [53, 54]). While at this level the physical mechanisms behind such results remain somewhat mysterious, the derivation from the gravitational path integral

[1] As an example, the firewall proposal of [23, 40, 41] did nothing to explain the dynamics from which the supposed firewall might arise.

nevertheless implies that the explicit addition of novel physics is not required. Indeed, it instead suggests that fundamental lessons might be revealed by carefully dissecting the relevant calculations and studying the path integral in more detail.

A starting point for such further investigation is the observation of [49] that the above replica trick results appear to be inconsistent with one might normally call a single well-defined theory. In particular, rather than taking single well-defined values, partition-function-like quantities seem to have both a mean value and a non-zero variance. This feature is associated with the fact that dominant saddles in the replica computations involve connected bulk spacetimes with *disconnected* asymptotically AdS boundaries. Such geometries have been termed spacetime wormholes, or Euclidean wormholes when the geometry is Euclidean.

This relation will be reviewed below, but is familiar from older discussions [55–58]. In particular, refs. [55–57] argued that spacetime wormholes require the gravitational Hilbert space to include spacetimes with compact Cauchy surfaces, and thus for which space at a moment of time has no asymptotically AdS boundary. This part of the gravitational Hilbert space was called the baby universe sector. Furthermore, it was argued that entanglement with this sector typically led the rest of the theory (here the asymptotically AdS sector) to act as if it were part of an ensemble of theories. However, a particular member of the ensemble could be chosen by selecting an appropriate baby universe state.

Our goal here is to combine the above ideas to better understand the ensembles associated with replica trick computations and to extract implications for particular members of such ensembles. We begin in section 2 by reviewing the connection between spacetime wormholes and ensemble-like properties, and by revisiting the baby universe ideas of [55–57]. In doing so, we incorporate features associated with a negative cosmological constant and asymptotically AdS boundaries. This both strengthens the results and allows a useful change in perspective. In particular, we avoid the use of 'third quantized perturbation theory' and emphasize that certain results follow exactly from any well-defined path integral. We also focus on the key role played by null states.

The output is a description of how (say, partition-function-like) quantities at asymptotically AdS boundaries have a spectrum of possible values determined by the gravitational path integral. Below, we focus on quantities $Z[\tilde{J}^*, J]$ that might be interpreted as computing the inner product of a state created by a source J on the past half of the Euclidean AdS boundary with another state created by a source $\tilde{J} = (\tilde{J}^*)^*$ on the future half of a Euclidean AdS boundary, where $*$ denotes CPT conjugation. However, the most general partition-function-like quantities allowed by our formalism include quantities that in a dual CFT would describe matrix elements of operators as well as e.g. $\mathrm{Tr}\,\rho^n$ for a wide variety of density matrices. The Rényi entropies of [44, 49] are

then functions of these quantities. In accordwith the original works [55–57], our analysis will show that one may generally describe such quantities as bring drawn from an ensemble of their possible values with the particular ensemble specified by the choice of baby universe state.

After describing this framework in section 2, section 3 introduces some simple toy models in which the gravitational path integral can be performed exactly including the full sum over possible topologies. The toy models are topological and involve finite-dimensional Hilbert spaces. An interesting feature of the models is that the dimension of the asymptotically AdS Hilbert space becomes a random variable Z, whose value can be *less* than the naive number k of independent states in the theory. For $k > Z$, consistency turns out to arise from an exact degeneracy in the inner product defined by the gravitational path integral. This degeneracy means that many a priori independent states differ by a null state, and so should be regarded as linearly dependent in the gravitational Hilbert space. Section 4 relates this degeneracy to diffeomorphism invariance, black holes, and the Page curve, arguing in particular that the replica computations of [48, 49] will imply a corresponding degeneracy in more general contexts. In section 5, we describe the approximation in which wormhole effects are small, analogous to the third quantised formalism of [57], and emphasise that the appearance of null states is associated with the failure of this approximation. We close with some summary and final discussion in section 6.

2 The gravitational path integral with spacetime wormholes

2.1 Path integrals and ensembles

We begin by describing a natural set of observables in any theory of gravity. For definiteness and convenience, we will assume locally AdS_{d+1} asymptotics. This is the context in which we have the most control and the clearest interpretation in terms of possible CFT duals.

Our theory will be defined by the path integral over a set of fields (including a metric) denoted collectively by Φ, with action $S[\Phi]$. Each boundary is associated with a set of admissible boundary conditions labelled by J, describing the behaviour of the fields $\Phi \sim J$ near the given boundary. In particular, J includes a d-dimensional boundary metric on a boundary manifold $\mathcal{M}$. We will focus on the case where the boundary metric has Euclidean signature, but Lorentzian or complex metrics are also allowed. We will generally take each $\mathcal{M}$ to be connected, and introduce disconnected boundary manifolds by specifying multiple such boundaries, each with its own J. However, there is no harm in letting $\mathcal{M}$ be disconnected, and the notation below remains

consistent. For each field other than the metric, J typically includes a function on the d-dimensional boundary $\mathcal{M}$ specifying an appropriate boundary condition for that field; e.g., it will typically specify what in the AdS/CFT context is known as the "non-normalisable part" of the field. In all cases, by $S[\Phi]$, we then mean the holographically renormalised action with boundary condition J.

Now, the gravitational path integral with asymptotically AdS boundary conditions specified by J is usually interpreted as computing a partition function $Z[J]$. This is particularly familiar in the AdS/CFT context [59, 60] where it gives the partition function of the dual CFT[2], but the identification of this quantity as a partition function in fact dates back to the first discussions of Euclidean approaches to black hole thermodynamics (see e.g. [61]). Motivated by this interpretation, with an eye toward the ideas of [55–57], and following [62], we introduce the following notation for the path integral defined by an asymptotic boundary with n connected components, each with an associated J_i:

$$\Big\langle Z[J_1]\cdots Z[J_n]\Big\rangle := \int_{\Phi\sim J} \mathcal{D}\Phi\, e^{-S[\Phi]} \tag{2.1}$$

This equation *defines* the left hand side as the path integral over all configurations with n asymptotic boundaries with boundary conditions specified by $J_1,\ldots,J_n$. The notation is chosen to be suggestive of a particular interpretation to be described below.

The presence of spacetime wormholes in the path integral now leads to a phenomenon which is very puzzling from the standard AdS/CFT point of view [58, 63] (see [55, 56] for earlier discussions of the asymptotically flat analogue in which S-matrix elements play the role of our partition functions). The path integral (2.1) does generally not factorize over disconnected boundaries:

$$\Big\langle Z[J_1]Z[J_2]\Big\rangle \neq \Big\langle Z[J_1]\Big\rangle\Big\langle Z[J_2]\Big\rangle. \tag{2.2}$$

The difference between right and left sides arises because the sum over topologies in the Euclidean path integral for $\Big\langle Z[J_1]Z[J_2]\Big\rangle$ not only yields terms of the form T_1T_2 for any pair T_1, T_2 of terms associated separately with $\Big\langle Z[J_1]\Big\rangle$ and $\Big\langle Z[J_2]\Big\rangle$, but also contains additional contributions from terms in which the two boundaries lie in the same connected component of the bulk manifold; see figure 1. We use the term spacetime wormhole, or sometimes Euclidean wormhole, to refer to any such connection.

[2]We emphasize, however, that we allow very general notions of 'sources' and thus very general notions of 'partition functions.' In particular, one may use sources to prepare initial and final states and to insert operators, so that one should be able to represent any matrix element of any operator in the dual CFT should as some $Z[J]$. In the same way, any Rényi entropy of any state that can be prepared by sources (and perhaps restricted to any region) should again be some $Z[J]$.

Note that spacetime wormholes are generally localized in both space and time, and thus differ qualitatively from spatial wormholes like the familiar Einstein-Rosen bridge that exist on every smooth Cauchy slice of the maximally extended Lorentz signature Schwarzschild spacetime.

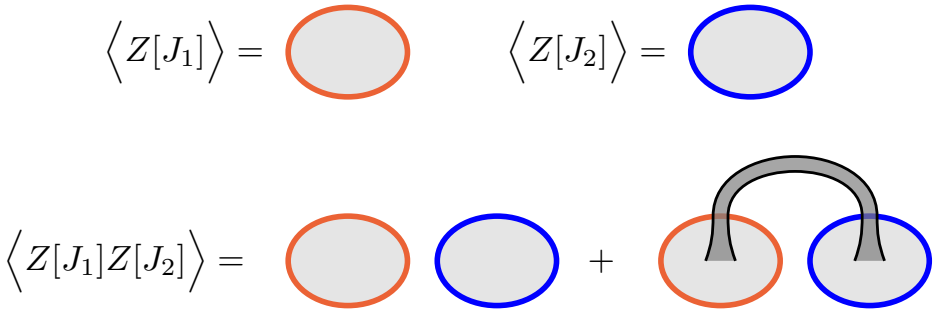


Figure 1: The gravitational path integral with spacetime wormholes does not factorize. The top line gives a diagramatic representation of the path integrals $\langle Z[J]_1 \rangle$ and $\langle Z[J_2] \rangle$ that would naively define partition functions $Z[J_1]$ and $Z[J_2]$. The natural path integral $\langle Z[J_1]Z[J_2] \rangle$ associated with a pair of boundaries yields all terms generated by multiplying $\langle Z[J_1] \rangle \langle Z[J_2] \rangle$, but also contains additional connected contributions schematically shown as the second term in the bottom line.

The two sides of (2.2) must thus differ unless the contributions with extra connections exactly cancel among themselves, or unless such contributions are excluded. The first option appears to require fine tuning, and the second the imposition of non-local constraints that undermine the presumed local nature of the theory. It is also difficult to see how one might introduce useful such constraints without destroying other apparent successes of the Euclidean path integral, such as the description of the Hawking-Page transition for AdS black holes, which is associated with a change in the topology of the dominant Euclidean saddle. We therefore allow terms with extra connections, and at least for the moment assume that they lead to a non-zero difference between the two sides of (2.2). It follows that we cannot simply interpret $\langle Z[J_1] \rangle$, $\langle Z[J_2] \rangle$ as partition functions with product $\langle Z[J_1]Z[J_2] \rangle$.

From the bulk point of view, the extra connections appear to describe dynamical interactions between a priori independent asymptotic regions. This point of view is not naturally compatible with standard AdS/CFT, but it may instead be consistent to interpret $\langle Z[J_1]Z[J_2] \cdots \rangle$ as the expectation value of a product of partition functions in an ensemble of boundary dual theories. In this interpretation, the connected contributions would describe probabilistic correlations from the ensemble average rather than dynamical interactions.

While these two interpretations may at first seem to be in tension, in analogous settings it was argued by [55–57] that they are in fact consistent. The rest of section 2 will be dedicated to providing a version of this discussion that incorporates features associated with asymptotically AdS boundaries. We find that using these new features allow strengthened conclusions, and perhaps as a result we will take a slightly different perspective than that of [55–57].

Before turning to the detailed discussion in section 2.2, it is useful to provide a brief overview. As in [55–57], the connection between the above two interpretations is motivated by realizing that summing over arbitrary topologies in our path integrals, and in particular over manifolds with arbitrary numbers of connected components, means that generic terms in $\langle Z[J_1]Z[J_2]\cdots\rangle$ contain factors associated with compact spacetimes having no boundaries whatsoever. The idea that the Hilbert space of a theory can be identified by cutting open the path integral then suggests that we should also slice open such compact spacetimes. Doing so identifies a new sector not associated on this slice with any of the asymptotically AdS boundaries, but which is instead associated with spatially compact universes; see figure 2. We call this the baby universe sector following [55–57], where the name comes from the idea that one can in many cases [64–67] think of the closed universe having been emitted by a (here asymptotically AdS) parent universe.

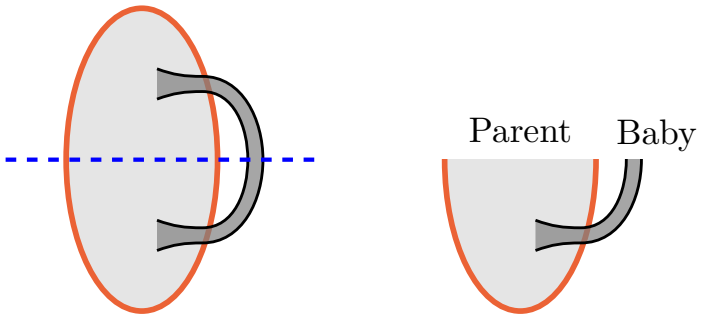


Figure 2: Slicing open a spacetime with a boundary and a handle (left) can give a disconnected geometry on the slice, including a closed 'baby universe' that has become detached from the parent asymptotically AdS universe. The baby universe does not intersect the asymptotically AdS boundary (red line) at the moment of time described by the indicated slice.

The discussion of baby universes is simplest in the context of Euclidean path integrals with boundary conditions J_i given by Euclidean metrics, but our discussion does not exclude more general contexts. In particular, one can choose boundary conditions with Lorentzian pieces of the metric, using a Schwinger-Keldysh type formalism

in which Euclidean sections of the metric are used to prepare states and Lorentzian sections give real time evolution. In such a case, it is useful to think of the gravitational path integral as involving complex metrics.

Such constructions allow us to describe quite general observables that might be associated with a putative dual CFT. Indeed, the set of observables we are using is also sufficient to describe coupling to an auxiliary quantum system, as is important in [43–48]. To do this, we can simply allow sources J to be operators in the auxiliary system, and then include a corresponding auxlliary path integral to compute the effects of such operators. We discuss this construction in more detail in section 4.

Note that the ability of Euclidean or complex universes to split and join as shown in figure 2 indicates that baby universes can affect the physics of universes with asymptotically AdS boundaries. In this context, it becomes clear that the definition of our path integral (2.1) includes an implicit choice of the initial and final state of closed baby universes. Most naturally, the path integral computes expectation values in the Hartle-Hawking no-boundary state [68], defined by the absence of additional boundaries besides those required by the $Z[J]$ insertions. But this is not the only choice of baby universe state that we can describe with our gravitational path integral, and other choices will be associated with different ensembles. In particular, we will construct special 'α-states' of baby universes in which the factorisation property is restored, and no ensemble is required.

One further comment is in order before turning to the details. In the above discussion we have written our amplitudes as if the path integral gives some definite, finite value. However, in all but the very simplest contexts, gravitational path integrals have been defined only as asymptotic expansions (perhaps with nonperturbative contributions) in some small coupling. Both loop expansions and sums over nonperturbative sectors will typically fail to converge, and there may be no obvious, natural or unique way to define a finite result. The distinction between exact quantities with finite values of parameters and asymptotic expansions may well be important, and we will return to this issue in section 6. Nonetheless, for the remainder of this section we will treat the path integral in (2.1) as if it gives well-defined exact results.

2.2 The baby universe Hilbert space

As described above, one can obtain a natural Hilbert space interpretation by cutting open the path integral (2.1). In particular, we split each history over which we sum into a 'past' and 'future' that meet on some slice where we imagine summing over a complete set of intermediate states. There is a choice of how we cut, constrained by the way in which the asymptotic boundaries are labelled past or future. For now, we will choose to place each connected component of the boundary either entirely to the

past or entirely entirely to future of our cut, so that our intermediate slice intersects no asymptotically AdS boundaries (generalizing in section 2.4). We thus identify the relevant Hilbert space as the space of closed universes in the theory. We call this the 'baby universe' Hilbert space $\mathcal{H}_{\rm BU}$ for the reasons described above.

One might hope to describe elements of the baby universe Hilbert space as wavefunctions of all possible spatial metrics (and field configurations on those metrics). A complication is that, as usual in a gravitational theory, diffeomorphism invariance forbids a notion of universal time that might be used specify precisely where the past/future cut is to be made. Proceeding in this manner would thus require imposing the gravitational constraints (the Wheeler-DeWitt equation) on the resulting wavefunctions. This is made particularly challenging in the current context where spacetime wormholes are important, so that the associated splitting and joining of universes should modify these constraints [57].

However, we can bypass these difficulties entirely by using our asymptotic boundaries to define states in the baby universe Hilbert space. Given a set $\{J_1, \ldots, J_m\}$ of boundary conditions, there is a state

$$\Big| Z[J_1] \cdots Z[J_m] \Big\rangle \in \mathcal{H}_{\rm BU}, \tag{2.3}$$

defined by the specified boundary conditions for the path integral. This is particularly natural for sources defining Euclidean signature boundary metrics and in the presence of a negative cosmological constant. While a negative cosmological constant tends to cause universes to collapse in Lorentzian time evolution (perhaps with a sinusoidal form), after Wick rotation to Euclidean signature it tends to cause accelerated expansion with respect to Euclidean time. As a result, such closed cosmologies naturally have Euclidean signature asymptotically AdS boundaries at infinite Euclidean times.

We will think of the boundary conditions associated with the state (2.3) as living 'in the past.' They can then be paired with bra-vectors living 'in the future' — though one should understand that these are simply names without intrinsic meaning. Note that the ordering of the $Z[J_i]$ in (2.3) is not important. Reordering the sources gives equivalent boundary conditions for the path integral, and so must define the same state. An important special case is $m = 0$, giving the Hartle-Hawking state with no boundary in the past:

$$\text{No boundaries } (m = 0) \longrightarrow \Big| \text{HH} \Big\rangle \in \mathcal{H}_{\rm BU}. \tag{2.4}$$

Here we emphasize that this is not just a state on a single universe, but that it instead represents a state of the full collection of an indefinite number of baby universes.

States of the form (2.3) defined by different sources, or even with different numbers of sources m, are generally not mutually orthogonal in any useful sense. Note that

the physical notion of inner product cannot simply be assumed to have any particular form, but is something we must compute from the theory. It must thus follow from an appropriate path integral. Now, some readers may be confused by the fact that in quantum field theory one typically uses first-quantized path integrals to compute Green's functions and not to compute inner products. However, as explained in e.g. [69], in defining the gravitational path integral one must make a choice — in some languages, associated with specifying the contour of integration — as to whether it fully imposes the gravitational constraints or instead defines a Green's function. We simply choose the former, and we take the correlators (2.1) to be computed with the same specifications. With this understanding, the path integral indeed computes the inner product[3] which is then given by

$$\Big\langle Z[\tilde{J}_1]\cdots Z[\tilde{J}_n]\Big| Z[J_1]\cdots Z[J_m]\Big\rangle = \Big\langle Z[\tilde{J}_1^*]\cdots Z[\tilde{J}_n^*]Z[J_1]\cdots Z[J_m]\Big\rangle. \tag{2.5}$$

Here the right hand side is just the amplitude defined in (2.1) with boundary conditions $Z[J]$ and $Z[\tilde{J}^*]$, and where $*$ is the CPT conjugate operation on boundary conditions J. This operation should have the property that if we act with $*$ on every boundary, the amplitude is complex conjugated:

$$\Big\langle Z[J_1^*]\cdots Z[J_n^*]\Big\rangle = \Big\langle Z[J_1]\cdots Z[J_n]\Big\rangle^*. \tag{2.6}$$

This guarantees that the inner product (2.5) is Hermitian. If we can interpret $Z[J]$ as random variables with correlation functions $\Big\langle Z[J_1]\cdots Z[J_n]\Big\rangle$, then (2.5) reduces to a standard construction in probability theory, in which the covariance matrix of pairs of random variables defines an inner product. In particular, showing that the amplitudes follow from expectation values of a distribution with nonnegative probabilities would imply that our inner product is positive semi-definite.

Note that the states (2.3) need not be normalised. In particular, the norm of the Hartle-Hawking state is given by what one might call the cosmological partition function $\mathfrak{Z}$, defined by the path integral over all spacetimes without boundary:

$$\mathfrak{Z} = \big\langle 1 \big\rangle = \big\langle \,\mathrm{HH}\,\big|\,\mathrm{HH}\,\big\rangle = \int_{\text{no boundary}} \mathcal{D}\Phi\, e^{-S[\Phi]}. \tag{2.7}$$

[3] In the language of Dirac constraint quantization [70], (2.5) corresponds to taking two arbitrary 'kinematic' states (which may not satisfy the constraints), projecting them onto the space of states satisfying the constraints, and computing the physical inner product of the resulting projections. See [71–75] for further comments, and [76–78] for connections to path integrals. As in [76, 77], using (2.5) corresponds to simply skipping to the final answer without going through the intermediate steps inherent in [70].

For most purposes, it would be sufficient to consider normalised amplitudes, where we divide by $\mathfrak{Z}$. This is equivalent to performing the path integral excluding closed components of spacetime which do not connect to any asymptotic boundary.

We now have a space of states defined by (finite) linear combinations of the states (2.3) in correspondence with formal polynomials of 'partition functions' $Z[J]$, and an inner product defined by extending (2.5) sesquilinearly. This is almost enough to construct a baby universe Hilbert space. The missing ingredient is a single property that we demand of our path integral (2.1), namely reflection positivity. This can be stated as the requirement that (2.5) defines a positive semidefinite inner product on finite linear combinations of states (2.3):

$$\big\|\Psi\big\|^2 := \big\langle\Psi|\Psi\big\rangle \geq 0 \text{ for all } \big|\Psi\big\rangle = \sum_{i=1}^{N} c_i \Big| Z[J_{i,1}]\cdots Z[J_{i,m_i}]\Big\rangle. \tag{2.8}$$

Thus is clearly required if our gravitational path integral is to define a standard quantum theory, though it is cumbersome to verify directly for all states. While this can be done for the simple toy models studied in section 3, for more complicated systems it would be very useful to find properties that imply (2.8) but are easier to check.

Assuming (2.8), we now *define* the baby universe Hilbert space $\mathcal{H}_{\rm BU}$ though a standard construction, as the completion of the space of linear combinations of states (2.3) with the inner product (2.5). Roughly speaking, states of $\mathcal{H}_{\rm BU}$ are infinite sums over states (2.3) with finite norm defined by (2.5).[4] Importantly, however, infinite sums with different terms and coefficients may not give rise to distinct states in $\mathcal{H}_{\rm BU}$. Equivalently, some infinite sums may be identified with the zero state in $\mathcal{H}_{\rm BU}$; i.e., for appropriate coefficients c_i one may find

$$\sum_{i=1}^{\infty} c_i \Big| Z[J_{i,1}]\cdots Z[J_{i,m_i}]\Big\rangle = 0. \tag{2.9}$$

Naively, the Hilbert space $\mathcal{H}_{\rm BU}$ may appear to consist of formal power series in the objects $Z[J]$ with some convergence property. But it is in fact smaller since the construction divides out by the set of 'null states' (2.9). This may seem like a minor technical point. Of course, from one perspective the inner product defined by any

[4] $\mathcal{H}_{\rm BU}$ is the set of equivalence classes of Cauchy sequences $\{|\Psi_i\rangle\}_{i\in\mathbb{N}}$, where two sequences $\{|\Psi_i\rangle\}$, $\{|\Phi_j\rangle\}$ are equivalent if $\| \, |\Psi_i\rangle - |\Phi_j\rangle \, \|^2 \to 0$ as $i,j\to\infty$. Recall that a sequence is Cauchy when $\| \, |\Psi_i\rangle - |\Psi_j\rangle \, \|^2 \to 0$ as $i,j\to\infty$. The inner product between two such sequences is defined by the limit of the inner products of the terms, which exists and is the same for all members of the equivalence class. $\mathcal{H}_{\rm BU}$ is then a Hilbert space, so in particular is complete and the inner product is positive definite. It is separable as long as the set of possible sources J has a countable dense subset (assuming that amplitudes are continuous in J).

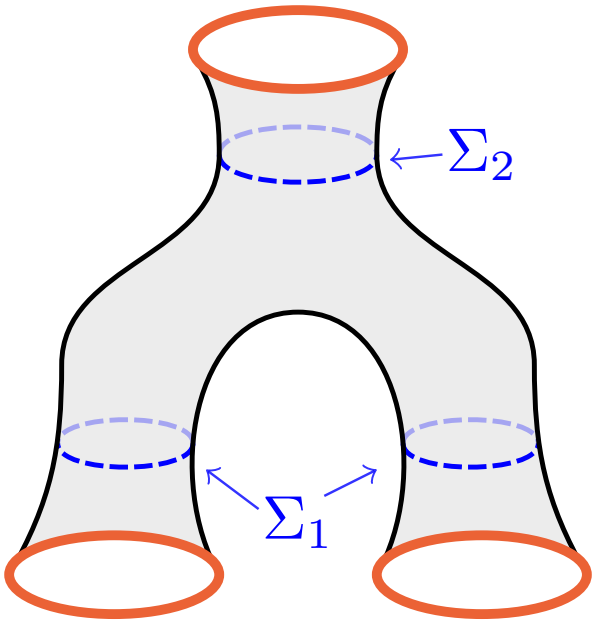


Figure 3: In the presence of spacetime wormholes, different spatial slices of a spacetime may have different number of connected components. Here, on the slice Σ_1 we have two circular universes, but on Σ_2 we have only one. These may be thought of as different gauge choices for the same state.

gravitational path integrals naturally leads to a large set of such null states due to the gravitational gauge symmetry. But we usually expect that symmetry to act trivially at the asymptotically AdS boundaries where our sources J are defined; i.e., natural sources J are invariant under familiar gravitational gauge symmetries. As a result, one might expect the null states to simply encode possible senses in which one may have accidentally introduced an overcomplete set of sources. However, one should expect the sum over topologies to modify the gravitational gauge invariance so that it no longer corresponds precisely to familiar diffeomorphisms. As illustrated in figure 3, one expects different slices of the same spacetime to describe gauge equivalent states. But including a sum over topologies means that two such slices may no longer be related by a diffeomorphism, and in fact that they need not even contain the same number of connected components for space at the given time. It will thus be important to compute the effects of this modified gauge symmetry rather than to assume that they take a familiar form. In particular, while one might naively expect the effect of such modifications to be small, we will find sections 3 and 4 that in certain circumstances they lead to dramatic physical consequences.

The above construction of $\mathcal{H}_{\rm BU}$ is very similar to the construction of the Hilbert

space of a quantum field theory from its correlation functions in the Wightman [79] or Osterwalder-Schrader (see **Theorem 3-7** of [80], [81]) reconstuction theorems. In this analogy, our objects $Z[J]$ correspond to (smeared) local operators inserted in the Euclidean past, and the inner products between states with finitely many operator insertions are given by the (Euclidean) Wightman functions. The Hilbert space is again defined by the above completion construction.

2.3 Operators and α-eigenstates

Having constructed the baby universe Hilbert space $\mathcal{H}_{\rm BU}$, we now introduce a set of operators acting on it. Here we once again find asymptotic boundaries useful. In particular, we take any boundary $Z[J]$ to define an operator $\widehat{Z[J]}$ on $\mathcal{H}_{\rm BU}$. The matrix elements of this operator are defined by a path integral over all configurations with boundaries specified by some initial and final states with an additional boundary $Z[J]$.

Since the labelling of boundaries as past, future, and in between does not affect the value of the path integral, the defining relation of the operator $\widehat{Z[J]}$ is

$$\Big\langle Z[\tilde{J}_1]\cdots Z[\tilde{J}_{\tilde{m}}]\Big|\widehat{Z[J]}\Big|Z[J_1]\cdots Z[J_m]\Big\rangle = \Big\langle Z[\tilde{J}_1]\cdots Z[\tilde{J}_{\tilde{m}}]\Big|Z[J]Z[J_1]\cdots Z[J_m]\Big\rangle. \tag{2.10}$$

Since the span of the bra-vectors in (2.10) is dense in $\mathcal{H}_{\rm BU}$, we may write the action of such operators as

$$\widehat{Z[J]}\Big|Z[J_1]\cdots Z[J_m]\Big\rangle = \Big|Z[J]Z[J_1]\cdots Z[J_m]\Big\rangle, \tag{2.11}$$

extending the action of such operators to the full Hilbert space $\mathcal{H}_{\rm BU}$ by continuity[5]. For later use, we note that (2.11) implies that our defining states may be created by acting with the $\widehat{Z[J]}$ operators on the Hartle-Hawking no-boundary state,

$$\Big|Z[J_1]\cdots Z[J_m]\Big\rangle = \widehat{Z[J_1]}\cdots\widehat{Z[J_m]}\Big|HH\Big\rangle, \tag{2.12}$$

and thus by combining (2.5) and (2.11) that we may identify our original path integral as computing correlators in $\Big|HH\Big\rangle$ as advertised earlier:

$$\Big\langle Z[J_1]\cdots Z[J_m]\Big\rangle = \Big\langle \mathrm{HH}\Big|\widehat{Z[J_1]}\cdots\widehat{Z[J_m]}\Big|\mathrm{HH}\Big\rangle. \tag{2.13}$$

We also see that the Hermitian conjugate of $\widehat{Z[J]}$ is given by taking the CPT conjugate of the source:

$$\widehat{Z[J]}^\dagger = \widehat{Z[J^*]} \tag{2.14}$$

[5]Strictly speaking, this is the case for bounded functions of the $Z[J_i]$. As usual, unbounded operators can be defined only on somewhat smaller domains.

Thus far, we have really defined the $\widehat{Z[J]}$ as operators on the baby universe pre-Hilbert space (before taking the quotient by null vectors (2.9)). To show that $\widehat{Z[J]}$ is well-defined on $\mathcal{H}_{\rm BU}$, we must show that it maps null states to null states. But this follows immediately from either (2.14) or (2.12). In particular, for any null state $\big|\mathcal{N}\big\rangle$ and an arbitrary state $\big|\Psi\big\rangle$, we may define $\big|\Psi'\big\rangle = \widehat{Z[J^*]}\big|\Psi\big\rangle$ to write

$$\big\langle\Psi\big|\widehat{Z[J]}\big|\mathcal{N}\big\rangle = \big\langle\Psi'\big|\mathcal{N}\big\rangle = 0. \tag{2.15}$$

The last equality follows from the fact that $\big|\mathcal{N}\big\rangle$ is null, and since $\big|\Psi\big\rangle$ is arbitrary we see that $\widehat{Z[J]}\big|\mathcal{N}\big\rangle$ is also null as desired.

The set of operators $\widehat{Z[J]}$ for all possible J turns out to have a powerful set of properties. Firstly, since the states $\big|Z[J_1]\cdots Z[J_m]\big\rangle$ are unchanged by permutations of the sources J_i, it follows immediately from (2.11) that all $\widehat{Z[J]}$ mutually commute[6]:

$$\Big[\widehat{Z[J]}, \widehat{Z[J']}\Big] = 0. \tag{2.16}$$

In particular, this implies that each $\widehat{Z[J]}$ is normal (that is, it commutes with its Hermitian conjugate), so that we may apply the spectral theorem. It then follows from (2.16) that the Hilbert space $\mathcal{H}_{\rm BU}$ has a basis of orthonormal states $|\alpha\rangle$ which are simultaneous eigenvectors for all $\widehat{Z[J]}$ operators:

$$\widehat{Z[J]}|\alpha\rangle = Z_\alpha[J]|\alpha\rangle \quad \forall J. \tag{2.17}$$

Following [55], we call these α-eigenstates, or α-states for short. The spectrum $\{Z_\alpha[J]\}_\alpha$ of $\widehat{Z[J]}$ may be either discrete or continuous. In the latter case the $|\alpha\rangle$ are not normalisable states, but are instead delta function normalized. However, for simplicity we use notation in either case as if $|\alpha\rangle$ are normalisable eigenvectors, writing

$$\langle\alpha'|\alpha\rangle = \delta_{\alpha'\alpha}\,, \tag{2.18}$$

leaving the appropriate modifications for continuous spectrum implicit.

It turns out that the set $\{\widehat{Z[J]}\}$ for all possible J in fact defines a *complete* commuting set of operators on $\mathcal{H}_{\rm BU}$, as the state $|\alpha\rangle$ is determined up to a phase by its eigenvalues $Z_\alpha[J]$. To see this, note that we can determine all matrix elements of $|\alpha\rangle$ via

$$\begin{aligned}\Big\langle Z[J_1]\cdots Z[J_n]\Big|\,\alpha\,\Big\rangle &= \Big\langle \mathrm{HH}\,\Big|\widehat{Z[J_1]}^\dagger\cdots\widehat{Z[J_n]}^\dagger\Big|\,\alpha\,\Big\rangle \\ &= Z_\alpha[J_1^*]\cdots Z_\alpha[J_n^*]\big\langle\,\mathrm{HH}\,\big|\,\alpha\,\big\rangle.\end{aligned} \tag{2.19}$$

[6] A similar result was derived in [55–57] using an additional assumption about locality of induced couplings. Crucially, this assumption played no role in our argument above.

This means that the α-states define a preferred orthonormal basis for $\mathcal{H}_{\rm BU}$; we can even fix phases by choosing $\langle \mathrm{HH}|\alpha\rangle > 0$.

The above calculation of the matrix elements also shows that the Hartle-Hawking state has non-zero overlap with every α-state, $\langle \mathrm{HH}|\alpha\rangle \neq 0$. Otherwise $|\alpha\rangle$ has vanishing overlap with a dense set of states, and hence must be the zero state. If we define p_α by these overlaps according to

$$p_\alpha = \frac{|\langle\, \mathrm{HH}\,|\,\alpha\,\rangle|^2}{\langle\, \mathrm{HH}\,|\,\mathrm{HH}\,\rangle}, \qquad . \tag{2.20}$$

we find

$$p_\alpha > 0, \quad \sum_\alpha p_\alpha = 1, \tag{2.21}$$

where the second follows from completeness and orthonormality of the α basis. Now, by inserting complete sets of α-states, we can compute the general amplitude (2.1):

$$\begin{aligned}\Big\langle Z[J_1]\cdots Z[J_n]\Big\rangle &= \sum_{\alpha_0,\alpha_1,\ldots,\alpha_n} \langle\, \mathrm{HH}\,|\alpha_0\rangle\langle\alpha_0|Z[J_1]|\alpha_1\rangle\cdots\langle\alpha_{n-1}|Z[J_n]|\alpha_n\rangle\langle\alpha_n|\,\mathrm{HH}\,\rangle \\ &= \mathfrak{Z}\ \sum_\alpha p_\alpha Z_\alpha[J_1]\cdots Z_\alpha[J_n].\end{aligned} \tag{2.22}$$

The normalising factor $\mathfrak{Z}$ is the norm of the Hartle-Hawking state (2.7).

Equation (2.22), along with (2.21), tells us that a gravitational path integral (2.1) is quite generally compatible with an ensemble interpretation, exemplified by the matrix ensemble dual to JT gravity in [62], and analogous to the random couplings of [55, 56]. Specifically, the parameters α label the various theories in the ensemble, the eigenvalues $Z_\alpha[J]$ give definite values for observables in the theory associated with the particular label α, and p_α gives the probability of selecting α from the ensemble. The states $|\alpha\rangle$ making up our preferred eigenbasis of $\mathcal{H}_{\rm BU}$ are in one-to-one correspondence with members of the ensemble. A less extreme example of α-states is provided by the 'eigenbranes' described in [82] in the context of JT gravity, which act to constrain the eigenvalues of $\widehat{Z[J]}$, thus partially diagonalizing these operators. Note that we arrived at a classical probability distribution because the relevant operators are mutually commuting (2.16). The only property required of the gravitational path integral (besides its existence) was reflection positivity, to guarantee nonnegative probabilities.

With our new Hilbert space point of view, it is now clear that the ensemble described above is not unique. Instead, through (2.13) it was associated with the implicit choice of the Hartle-Hawking state in $\mathcal{H}_{\rm BU}$. While the Hartle-Hawking state is a particularly simple and natural choice, we are nevertheless free to select any state we like.

In particular, if the initial state of the baby universes is an α-state, this selects a single member of the ensemble so that amplitudes factorize:

$$\langle\alpha|\widehat{Z[J_1]}\widehat{Z[J_2]}|\alpha\rangle = \langle\alpha|\widehat{Z[J_1]}|\alpha\rangle\langle\alpha|\widehat{Z[J_2]}|\alpha\rangle = Z_\alpha[J_1]Z_\alpha[J_2]. \tag{2.23}$$

Any other state $|\Psi\rangle$ is a superposition of α-states, and describes an ensemble with probabilities $p_\alpha = |\langle\alpha|\Psi\rangle|^2$. Classical probabilities are sufficient to describe the ensemble, since relative phases between different α-states in the superposition are irrelevant for correlation functions of the commuting operators $\widehat{Z[J]}$. In other words, with respect to the algebra of the $\widehat{Z[J]}$, the α-states define superselection sectors.

If the path integral (2.1) already defines factorising amplitudes, so that our theory of gravity has a single boundary dual, we have a trivial special case of the formalism described here. In that case, the operators $\widehat{Z[J]}$ are constants $Z[J]$, and the Hilbert space of closed universes $\mathcal{H}_{\rm BU}$ is one-dimensional, spanned by the Hartle-Hawking state, which is also the unique α-state. We discuss this possibility further in section 6.

2.4 More Hilbert spaces

The above discussion concerned the Hilbert space $\mathcal{H}_{\rm BU}$ of closed 'baby' universes. We constructed $\mathcal{H}_{\rm BU}$ by cutting amplitudes in such a way that any given asymptotic boundary lies completely on one side of the cut. We now generalize this construction to allow cuts that intersect one or more components of the asymptotic boundary, thus splitting such boundary components into two parts. This gives us many different Hilbert spaces depending on the boundary conditions at the intersection, and in particular on the choice of a $(d-1)$-dimensional (perhaps oriented) spatial boundary geometry Σ. We thus call the resulting Hilbert space $\mathcal{H}_\Sigma$, leaving implicit the other sources J on Σ. Note that Σ can have any number of connected components, and if Σ is empty we find again the Hilbert space $\mathcal{H}_{\Sigma=\varnothing} = \mathcal{H}_{\rm BU}$ of closed baby universes described above.

The construction of $\mathcal{H}_\Sigma$ proceeds much as for $\mathcal{H}_{\rm BU}$, except that in addition to closed asymptotic boundary conditions denoted by $Z[J]$ we also have objects $\psi[J]$ defining boundary conditions on a piece $\mathcal{M}$ of an asymptotic boundary with $\partial\mathcal{M} = \Sigma$. As before, the manifold $\mathcal{M}$, and in particular its boundary Σ, is implicitly included in the sources J. For example, in the right panel of figure 2, $\mathcal{M}$ is the solid black semicircle forming the past asymptotically AdS boundary and Σ consists of the right and left endpoints. In a dual interpretation, $\psi[J]$ would define a state on the CFT Hilbert space with spatial geometry Σ, as the wavefunction for a given CFT field configuration on Σ would be computed by a path integral on $\mathcal{M}$ with sources J.

As before, we may choose $\mathcal{M}$ to be connected. Note that this does not imply $\Sigma = \partial\mathcal{M}$ to be connected. When Σ is not, it can be useful to write Σ as the disjoint

union $\Sigma = \Sigma_1 \sqcup \cdots \sqcup \Sigma_m$ of components Σ_i (where the ordering of the components is meaningful, in case they have the same geometry). Generalizing (2.3), we then have states

$$\Big|\psi[J_1]\cdots\psi[J_m]Z[J'_1]\cdots Z[J'_n]\Big\rangle \in \mathcal{H}_\Sigma, \tag{2.24}$$

where $\psi[J_i]$ is associated with component Σ_i for any source J. While this notation is useful, it is also somewhat awkward if we take a given $\psi[J_i]$ to be associated with a connected $\mathcal{M}_i$, whose boundary $\partial\mathcal{M}_i = \Sigma_i$ may again be disconnected. As a result, one will sometimes need to use a number of distinct decompositions $\Sigma = \Sigma_1 \sqcup \cdots \sqcup \Sigma_m$ (perhaps with different values of m) for a given $\mathcal{H}_\Sigma$.

The inner product on $\mathcal{H}_\Sigma$ generalizes (2.5) in a natural way if we note that a boundary condition $\psi[\tilde{J}_i]$ in the 'bra' (on some $\tilde{\mathcal{M}}_i$ with $\partial\tilde{\mathcal{M}}_i = \Sigma_i$) can be paired with a boundary condition $\psi[J_i]$ in the 'ket' (again on some $\mathcal{M}_i$ with $\partial\mathcal{M}_i = \Sigma_i$) to define a boundary condition $Z[\tilde{J}^*, J]$ associated with the closed boundary manifold $\tilde{\mathcal{M}}_i^*\mathcal{M}_i$ constructed by taking the manifold $\tilde{\mathcal{M}}_i^*$ (formed from $\tilde{\mathcal{M}}_i$ by reversing the orientation) and sewing $\tilde{\mathcal{M}}_i^*$ to $\mathcal{M}_i$ along Σ_i. In $Z[\tilde{J}^*, J]$, $*$ again denotes CPT conjugation of sources, and the sources on $\tilde{\mathcal{M}}^*\mathcal{M}$ are given locally by $\tilde{J}^*, J$. One may also wish to restrict the allowed sources to vanish sufficiently quickly at Σ_i so that the sources defined on $\tilde{\mathcal{M}}_i\mathcal{M}_i$ by such sewings are sufficiently smooth.

It is important that the above sewing is uniquely defined even when Σ_i admits isometries. In particular, recall that the above discussion fixed a manifold $\Sigma \supseteq \Sigma_i$ from the beginning, and at no point was there a quotient by diffeomorphisms of Σ. The individual points of Σ should thus be thought of as carrying definite labels, defining the unique sewing of $\tilde{\mathcal{M}}$ to $\mathcal{M}$. In particular, the notation in (2.24) is not invariant under reordering of the Σ_i.

We shall write the pairing as $Z[\tilde{J}^*, J] = \Big(\psi[\tilde{J}], \psi[J]\Big)$. This notation is chosen be suggestive of an inner product $(\cdot,\cdot)$ of states in the dual CFT Hilbert space. The distinguishability of points in Σ is motivated either by a dual CFT perspective, or from familiar gravitational boundary conditions at asymptotically AdS boundaries. The extended inner product is then defined by using the above pairing and and evaluating the resulting path integral as before:

$$\Big\langle \psi[\tilde{J}]\Big|\psi[J]\Big\rangle = \Big\langle \Big(\psi[\tilde{J}], \psi[J]\Big)\Big\rangle = \Big\langle Z[\tilde{J}^*, J]\Big\rangle \tag{2.25}$$

We emphasize again that if Σ contains identical connected components Σ_1, Σ_2, the components are treated as distinguished and canonically ordered. Thus in the notation of (2.24), $\big|\psi[J_1]\psi[J_2]\big\rangle \neq \big|\psi[J_2]\psi[J_1]\big\rangle$. While the norms of these states will agree, the inner product of these states with generic other kets will not (for example, $\big\langle\psi[J_2]\psi[J_1]\big|\psi[J_1]\psi[J_2]\big\rangle = \big\langle Z[J_2^*, J_1]Z[J_1^*, J_2]\big\rangle \neq \big\langle Z[J_2^*, J_2]Z[J_1^*, J_1]\big\rangle$, even if $\Sigma_1 = \Sigma_2$

so this pairing makes sense). This is a special case of the statement that states need not be invariant under symmetries of Σ.

As in the discussion of $\mathcal{H}_{\rm BU}$, the structure above is properly described as being pre-Hilbert space. The actual Hilbert space $\mathcal{H}_\Sigma$ is then constructed as a completion, which includes a quotient with respect to the space of null vectors. This procedure succeeds when the path integral is appropriately reflection positive, by which we mean that the inner product it defines on the pre-Hilbert space is positive semi-definite. The inner product on the final $\mathcal{H}_\Sigma$ is then positive definite as desired. Note that reflection positivity on $\mathcal{H}_\Sigma$ is an additional requirement we impose on the path integral, not necessarily implied by reflection positivity on $\mathcal{H}_{\rm BU}$; this will prove to be relevant for the toy model discussed in section 3.

As before, we have operators $\widehat{Z[J]}$ acting on the Hilbert spaces $\mathcal{H}_\Sigma$, and in particular which preserve the space of null states in the pre-Hilbert space for the same reason as before. Again, these operators mutually commute. But now we also have a plethora of new operators which can map between Hilbert spaces with different boundaries. In particular, if $\psi[J]$ is associated with $\mathcal{M}$ having $\partial\mathcal{M} = \Sigma$, then for any $\tilde{\Sigma}$ there is an operator

$$\widehat{\psi[J]} : \mathcal{H}_{\tilde{\Sigma}} \to \mathcal{H}_{\Sigma\sqcup\tilde{\Sigma}}, \tag{2.26}$$

$$\text{with} \quad \widehat{\psi[J]}\Big|\tilde{\psi}[\tilde{J}]Z[J_1']\cdots Z[J_n']\Big\rangle = \Big|\psi[J]\tilde{\psi}[\tilde{J}]Z[J_1']\cdots Z[J_n']\Big\rangle, \tag{2.27}$$

where in $\Sigma\sqcup\tilde{\Sigma}$ we define the components of Σ to be ordered before components of $\tilde{\Sigma}$. We may use (2.26) even when $\mathcal{M}, \tilde{\mathcal{M}}$ are disconnected. Note, however, that (when $\Sigma \neq \Sigma'$) it does not make sense to ask whether $\widehat{\psi[J]}, \widehat{\psi[J']}$ commute, as $\widehat{\psi[J]}\widehat{\psi[J']}$ maps $\mathcal{H}_{\tilde{\Sigma}} \to \mathcal{H}_{\Sigma\sqcup\Sigma'\sqcup\tilde{\Sigma}}$, while $\widehat{\psi[J']}\widehat{\psi[J]}$ maps $\mathcal{H}_{\tilde{\Sigma}} \to \mathcal{H}_{\Sigma'\sqcup\Sigma\sqcup\tilde{\Sigma}}$.

Nevertheless, one can build a dense set of states in $\mathcal{H}_\Sigma$ by acting with such operators on $\mathcal{H}_\varnothing = \mathcal{H}_{\rm BU}$. As a result, the fact that $\widehat{\psi[J]}$ preserves the null space, and thus is truly well-defined on $\mathcal{H}_\Sigma$, follows from (2.25) and the corresponding property for $\widehat{Z[\tilde{J}^*, J]}$.

The adjoint operator $\widehat{\psi[J]}^\dagger$ maps from $\mathcal{H}_{\tilde{\Sigma}\sqcup\Sigma}$ to $\mathcal{H}_{\tilde{\Sigma}}$ by taking the boundary conditions defined by the state on which it acts, and gluing to boundary conditions of the CPT conjugate source J^* along the manifold Σ.

Since the $\widehat{Z[J]}$ commute, it is again useful to diagonalize them using α-states. Thus the Hilbert space splits as

$$\mathcal{H}_\Sigma = \bigoplus_\alpha \mathcal{H}_\Sigma^\alpha. \tag{2.28}$$

One can explicitly build the spaces $\mathcal{H}_\Sigma^\alpha$ from the α-states of $\mathcal{H}_{\rm BU}$, as we may define

$$\Big|\psi[J_1]\cdots\psi[J_m];\alpha\Big\rangle := \widehat{\psi[J_1]}\cdots\widehat{\psi[J_m]}\Big|\alpha\Big\rangle \in \mathcal{H}_\Sigma^\alpha, \tag{2.29}$$

and, the states (2.29) are dense in $\mathcal{H}^\alpha_\Sigma$. In the special case $\Sigma = \varnothing$ corresponding to $\mathcal{H}_{\rm BU}$, each $\mathcal{H}^\alpha_\varnothing$ is one dimensional, consisting of multiples of $|\alpha\rangle$. It follows that all of our boundary operators leave α unchanged. For example, evaluating the analogue of (2.25) in α-states we have

$$\big\langle \psi[J_2];\alpha_2 \big| \psi[J_1];\alpha_1 \big\rangle = Z_{\alpha_1}[J_2^*, J_1]\, \delta_{\alpha_1\alpha_2}. \tag{2.30}$$

It also follows that $\widehat{Z[J]}$ commutes with $\widehat{\psi[\tilde{J}]}$.

Finally, note that there is a natural map Υ from $\mathcal{H}_{\Sigma_1} \otimes \mathcal{H}_{\Sigma_2}$ into $\mathcal{H}_{\Sigma_1 \sqcup \Sigma_2}$ defined by concatenation of sources:

$$\begin{aligned} &\big|\psi[J_{11}]\cdots\psi[J_{1,m_{\Sigma_1}}]Z[J'_{11}]\cdots Z[J'_{1,n_1}]\big\rangle \otimes \big|\psi[J_{21}]\cdots\psi[J_{2,m_{\Sigma_2}}]Z[J'_{21}]\cdots Z[J'_{2,n_2}]\big\rangle \\ &\mapsto \big|\psi[J_{11}]\cdots\psi[J_{1,m_{\Sigma_1}}]\psi[J_{21}]\cdots\psi[J_{2,m_{\Sigma_2}}]Z[J'_{11}]\cdots Z[J'_{1,n_1}]Z[J'_{21}]\cdots Z[J'_{2,n_2}]\big\rangle. \end{aligned} \tag{2.31}$$

This maps acts nicely within each α-sector, taking $\mathcal{H}^\alpha_{\Sigma_1} \otimes \mathcal{H}^\alpha_{\Sigma_2}$ into $\mathcal{H}^\alpha_{\Sigma_1\sqcup\Sigma_2}$. In particular, since acting on $|\mathrm{HH}\rangle$ with the $\widehat{Z[J]}$ yields a dense set of states in $\mathcal{H}_{\rm BU}$, one may write $|\alpha\rangle = f_\alpha(\{Z[J_i]\})|\mathrm{HH}\rangle$ for some function f_α that takes the value 1 on arguments $\{Z_\alpha[J_i]\}$ but which vanishes on $\{Z_{\alpha'}[J_i]\}$ for all $\alpha' \neq \alpha$. One then finds

$$\big|\alpha\big\rangle \otimes \big|\alpha'\big\rangle = f_\alpha f_{\alpha'} \big|\mathrm{HH}\big\rangle = \delta_{\alpha,\alpha'}\big|\alpha\big\rangle, \tag{2.32}$$

and more generally

$$\Upsilon : \mathcal{H}^\alpha_{\Sigma_1} \otimes \mathcal{H}^{\alpha'}_{\Sigma_2} \to \delta_{\alpha,\alpha'}\mathcal{H}^\alpha_{\Sigma_1\sqcup\Sigma_2}. \tag{2.33}$$

Here we have used the notation $c\mathcal{H}$ for non-negative real c to denote a Hilbert space with inner product c times that of $\mathcal{H}$. In particular, $c\mathcal{H} = \{0\}$ for $c = 0$. We will use Υ_α to denote the restriction of Υ to diagonal tensor products of the form $\mathcal{H}^\alpha_{\Sigma_1} \otimes \mathcal{H}^\alpha_{\Sigma_2}$.

It is natural to attempt to interpret $\mathcal{H}^\alpha_\Sigma$ as the Hilbert space of a dual CFT $\mathcal{C}_\alpha$ on Σ; this is the natural formulation of an isomorphism between bulk and boundary Hilbert spaces in the context of ensembles and baby universes. In this case, we would expect Υ_α to be an isomorphism, since this property would certainly hold true in a local dual theory. But this is not always the case, as the map may not be surjective; we will discuss an explicit example in section 3.6. The failure of Υ_α to be an isomorphism is a precise version of another potential 'factorisation problem' [83–85], which differs from the partition function factorisation problem discussed in the introduction and the start of this section. This new issue is naturally associated with spatial wormholes while (2.2) is related to spacetime wormholes. In particular, the factorization problem of [83–85] occurs when there are two-sided black hole states with a spatial wormhole (Einstein-Rosen bridge) which cannot be represented as superpositions of products

of 'microstates' in the corresponding one-sided Hilbert spaces. For example, in a bulk theory with a standard Maxwell field but no charged particles, there are eternal charged black holes but no one-sided counterparts. An extreme version appears in pure JT gravity, which has a two-boundary Hilbert space but no single-sided Hilbert space. We expect that this feature is an artefact of simple toy models, and would be absent in more realistic theories.

3 Example: a very simple topological theory

This section further explores the structure described in section 2 in very simple theories of two-dimensional gravity. Indeed, the model described in section 3.1 is plausibly the simplest possible such theory. Our models are inspired by recent work studying spacetimes of nontrivial topology in JT gravity [62, 86, 87], along with the addition of 'end-of-the-world brane' dynamical boundaries [49]. We further simplify that class of models by removing any notion of a dynamical metric or dilaton, leaving a theory of topology alone. The resulting models are tractable enough to be solved exactly, and for many details to be made explicit. They thus give a surprisingly clean illustration of the ideas of section 2, and demonstrate the type of results to which such ideas can lead.

We begin by presenting the simplest model (without end-of-the-world branes) in section 3.1. This theory allows only one boundary condition Z, associated with a single operator $\widehat{Z}$ of the class described in section 2.3, with the path integral defined by a single bulk parameter S_0 determining the suppression of nontrivial topology, along with a (somewhat ad hoc) parameter S_∂ associated with boundaries, whose preferred value $S_\partial = S_0$ will be determined later by a consistency analysis in section 3.7. We then evaluate its amplitudes in section 3.2 and construct the Hilbert space of closed universes $\mathcal{H}_{\rm BU}$ in section 3.3. The most interesting output of this model is that the spectrum of $\widehat{Z}$ turns out to be non-negative and discrete, and in fact takes non-negative integer values for $S_\partial = S_0$, compatible with an interpretation as the dimension of a dual Hilbert space. The model with end-of-the-world branes is then described in section 3.4, and its α-states are described in section 3.5. Here we find that, no matter how many species k of end-of-the-world brane states we allow, for $S_\partial = S_0$ all α-states define an inner product on end-of-the-world brane states with rank equal to or less than the eigenvalue Z_α of $\widehat{Z}$, compatible with states in a dual Hilbert space of dimension Z_α. This remarkable compression of the Hilbert space illustrates the importance of understanding the null states 2.9 in extracting the correct physics. It also shows in this model that results analogous to the Rényi entropy computations of [48, 49] will hold not

just for typical members of the ensemble defined by the Hartle-Hawking no-boundary state, but in fact for all allowed α-states.

We then return to the ad hoc parameter S_∂ in section 3.7. First, we describe how different choices for this parameter modify the model. We find that for generic S_∂ (and in particular $S_\partial = 0$) the end-of-the-world brane models fail to be reflection positive, and find the set of S_∂ for which reflection positivity holds true. For values of S_∂ satisfying reflection positivity for any number k of end-of-the-world brane states, the spectrum of $\widehat{Z}$ is a subset of the non-negative integers and the rank of the end-of-the-world brane Hilbert space is bounded as above. In particular, the reflection positive models have all the properties required to interpret Z_α as the dimension of a Hilbert space which contains the end-of-the-world brane states.

3.1 A theory of topological surfaces

We now consider a theory of purely topological two-dimensional gravity in which spacetime is a two-dimensional manifold[7] (surface), but the only additional structure we introduce is an orientation. We thus have neither a spacetime metric nor the conformal or complex structure that would appear in the standard model of topological gravity [88]. The histories that can appear in a path integral are then the set of oriented topological surfaces with boundaries dictated by the relevant boundary conditions. This set is discrete and (for each connected component) is famously classified by genus and number of circular boundaries [89, 90]. Since there is no possibility to add sources in this model, we simply use Z to denote the boundary condition on any circular boundary.[8]

In this first model, the only boundaries are those fixed by boundary conditions. As described in section 2.4, such boundaries should be thought of as distinguishable even when their boundary conditions coincide. As a result, the space of allowed configurations is the set of oriented surfaces with *labelled* boundaries, and two such configurations are considered equivalent only when they are related by a diffeomorphism that preserves each boundary separately.

We therefore define our path integral as a sum over such diffeomorphism classes of surface M. Nevertheless, residual effects of diffeomorphism invariance can lead to a nontrivial measure $\mu(M)$ on this space. This can arise when a group $\Gamma(M)$ of residual gauge symmetries remains after gauge fixing diffeomorphisms. This naturally leads to symmetry factors in the measure, of the form $\mu(M) = \frac{1}{|\Gamma(M)|}$. One may therefore expect

[7] For definiteness, we take smooth (not just topological) manifolds, and accordingly use the language of equivalence under diffeomorphisms rather than homeomorphisms.

[8] We take the set of boundary conditions to be a vector space, so that a general boundary condition assigns a (perhaps complex) weight to each non-negative integer n enumerating the possible numbers of circular boundaries.

to write our path integral in the tentative form

$$\int \mathcal{D}\Phi\, e^{-S[\Phi]} := \sum_{\text{Surfaces } M} \mu(M)\, e^{-S[M]}, \tag{3.1}$$

where we sum over surfaces M obeying the appropriate boundary conditions, up to diffeomorphisms acting trivially on the boundaries, weighted by an action $S[M]$.

One would ideally like to derive the measure factor $\mu(M)$ from a more complete model. Here, we will be content to define the model with a well-motivated choice of measure that leads to natural results. Since boundaries are distinguishable, and since any two surfaces related by boundary-preserving diffeomorphisms are already considered equivalent, we will assume the trivial measure $\mu(M) = 1$ for any connected manifold. It then remains to discuss only contributions to $\mu(M)$ from boundary-preserving diffeomorphisms that interchange the connected components of M. These can act only on compact connected components (i.e., the ones that have no boundary). With this understanding, the detailed form of $\mu(M)$ turns out to have little effect on the physics of interest. It leads only to a change of the 'cosmological partition functon' $\mathfrak{Z}$, the sum over compact universes, which is an overall normalisation of amplitudes (though at the end of section 3.3 we will encounter a situation in which our choice of measure is physically important). Nevertheless, we regard diffeomorphisms that permute compact connected components (necessarily with the same genus g) as residual gauge symmetries, and divide by the number of such permutations in the measure. This means that, if M has m_g connected components of genus g with no boundary for each g, we have

$$\mu(M) = \frac{1}{\prod_g m_g!}. \tag{3.2}$$

Following the principles of effective field theory, we should now write down the most general action allowed by the degrees of freedom. Fortunately, with only the topological degrees of freedom available to us, there is a unique local such action $S(M) = -S_0\chi(M)$, proportional to the Euler characteristic χ of spacetime[9], with a unique free parameter S_0. This is the Einstein-Hilbert action in two dimensions, and is the topological term of the action in JT gravity.

Despite the apparent uniqueness for the action, we now introduce an additional term $-S_\partial|\partial M|$, where $|\partial M|$ denotes the number of circular boundaries of M. As forewarned in the introduction to this section, for the moment the extra parameter

[9]Here we take locality to mean invariance under cutting and gluing surfaces. A precise version of the above statement is then that $\exp(S_0\chi)$ is the most general form for the amplitudes of a two-dimensional topological quantum field theory (TQFT) with trivial (one-dimensional) Hilbert space on the circle.

S_∂ appears completely ad hoc. In particular, while this is an intrinsic function of asymptotic boundaries, it is not a *local* counterterm. Indeed, as stated above, we expect that the unique local theory of our form is given by setting $S_\partial = 0$. We discuss how this factor may arise in 3.7 below, perhaps most simply by introducing a new local degree of freedom residing on boundaries. For now we simply note that the parameter effectively just rescales the definition of Z; i.e., it can be removed by introducing $\tilde{Z} = e^{S_\partial} Z$ and replacing each Z in (3.1) by $\tilde{Z}$.

Since all values of S_∂ are related by this scaling, it suffices to discuss only a single value in detail, and then to use the above scaling to understand all other values. Until section 3.7, we will thus confine discussion to the particularly simple case $S_\partial = S_0$. As an a posteriori justification, we will show in section 3.7 that the end-of-the world brane models fail to be reflection positive when $S_\partial = 0$, and $S_\partial = S_0$ is the most natural choice to cure this failure.

Our action is thus given by

$$S(M) = -S_0 \chi(M) - S_\partial \, n(M), \tag{3.3}$$

$$\text{where we choose } S_\partial = S_0 \quad \text{(until section 3.7)}. \tag{3.4}$$

The practical simplification of choosing $S_\partial = S_0$ is that it precisely cancels boundary contributions to χ in the action. The amplitudes in our path integral thus take the form

$$\left\langle Z^n \right\rangle \quad = \sum_{\substack{M \text{ with} \\ |\partial M| = n}} \mu(M) e^{S_0 \tilde{\chi}(M)} \ , \tag{3.5}$$

which we have written in terms of a modified Euler characteristic $\tilde{\chi}$ that does not count boundaries and which is given simply by

$$\tilde{\chi} = \sum_{\substack{\text{Connected} \\ \text{components}}} (2 - 2g). \tag{3.6}$$

Here g is the usual genus of each connected component that counts handles.

It will be useful below to sometimes use an alternate presentation of the sum (3.5). Instead of summing over surfaces with labeled boundaries, we can write $\left\langle Z^n \right\rangle$ as a sum over ordered lists M_L of *connected manifolds*, and also where we choose not to label the boundaries. The number of ways to label the boundaries is then accounted for by including a separate factor of the multinomial coefficient $\frac{n!}{\prod_i n_i!}$, where n_i is the number of boundaries in the ith entry of the list M_L. As is well known, $\frac{n!}{\prod_i n_i!}$ gives precisely the number of ways to arrange n boundaries into lists of subsets that have n_i boundaries in the ith subset. For a list of length m, including a factor of $\frac{1}{m!}$ then accounts for

the fact that the components are not ordered in the original sum (3.5), and also for the factor of $\mu(M)$ that arises when some items in the list both coincide and have no boundaries (so that exchanging these items neither generates a new term in (3.5) nor generates a new partition of the n boundaries). Thus we may rewrite (3.5) as

$$\left\langle Z^n \right\rangle = \sum_{\substack{\text{Ordered lists } M_L \\ \text{of connected surfaces} \\ \text{with } n \text{ boundaries}}} \frac{n!}{m! \prod_i n_i!} e^{S_0 \tilde{\chi}} \ , \tag{3.7}$$

where n, m, and n_i are as above.

Before computing the amplitudes (3.5), it is useful to comment further on the interpretation of Z in terms of a putative dual $0+1$-dimensional quantum mechanics (which we will sometimes call a CFT in analogy with AdS/CFT). Each Z would be naturally associated with the path integral of this quantum mechanics on the circle, which would describe the partition function $\operatorname{Tr} e^{-\beta H}$ for a circle of length β. But since we have no metric, there is no notion of boundary length β, and invariance under diffeomorphisms of the boundary implies a vanishing Hamiltonian $H = 0$. This means we have a topological quantum mechanics (a one-dimensional TQFT) where the only observable is the trace of the identity operator, which is the dimension of the Hilbert space:

$$Z \stackrel{?}{=} \operatorname{Tr}_{\mathcal{H}_{\text{CFT}}} 1 = \dim \mathcal{H}_{\text{CFT}} \tag{3.8}$$

A unitary dual quantum mechanics is therefore characterised by Z taking a value in the natural numbers $\mathbb{N}$ (or perhaps by Z being infinite). In the presence of spacetime wormholes connecting these boundaries, it would thus seem natural to find that Z is a random variable taking nonnegative integer values. We will see below that this is precisely the case for our model.

3.2 Evaluating the amplitudes

We now solve for the amplitudes $\left\langle Z^n \right\rangle$ defined above. We begin by computing the no-boundary partition function $\mathfrak{Z}$ as in equation (2.7). This is the case $n = 0$, given by the sum over arbitrary compact spacetimes without boundary. For this, we first compute the sum λ over *connected* compact surfaces, which are classified by genus. The measure is trivial for a connected surface, i.e. $\mu(M) = 1$, so we have

$$\lambda := \sum_{\substack{\text{Connected} \\ \text{compact surfaces}}} e^{S_0 \chi} = \sum_{g=0}^{\infty} e^{S_0(2-2g)} = \frac{e^{2S_0}}{1 - e^{-2S_0}}. \tag{3.9}$$

With our amplitudes defined by (3.5), and in particular excluding boundaries from the count in the Euler character, the value of λ is always the amplitude for any con-

nected component of spacetime (with fixed but arbitrary boundaries) after summing over connected topologies. This property determines all amplitudes of the model.

In the usual way, one may write $\mathfrak{Z}$ as the exponential of the sum λ over connected surfaces. For this, it is important that we include symmetry factors in our definition of the measure $\mu(M)$. Indeed, the exponentiation is particularly explicit by using (3.7) with $n = n_i = 0$, in which lists of length m contribute $\frac{1}{m!}$ times the mth power of the sum in (3.9). We thus find

$$\mathfrak{Z} = \big\langle 1 \big\rangle = e^{\lambda} \,. \tag{3.10}$$

In particular, in our model the path integral defined by the sum over topologies converges.

We now introduce boundaries. To evaluate $\langle Z^n \rangle$, it is simplest to compute a generating function

$$\big\langle e^{uZ} \big\rangle = \sum_{n=0}^{\infty} \frac{u^n}{n!} \big\langle Z^n \big\rangle, \tag{3.11}$$

and to extract the amplitudes from a power series in the 'chemical potential' u. Again, we wish to write (3.11) as the exponential of a sum over connected geometries. This is precisely the usual combinatorics familiar from Feynman diagrams, but it can also be seen explicitly from (3.7) which gives

$$\big\langle e^{uZ} \big\rangle = \sum_{\substack{\text{Ordered lists } M_L \\ \text{of connected surfaces}}} \frac{u^{\sum_i n_i}}{m! \prod_i n_i!} e^{S_0 \tilde{\chi}(M_L)} \quad , \tag{3.12}$$

where m is the number of surfaces in the list M_L, and n_i for $i = 1, \ldots m$ is the number of boundaries of the ith surface in the list. Since $\tilde{\chi}$ for the disconnected surface M_L is the sum of $\tilde{\chi}$ for the individual components, this disconnected pieces exponentiate,

$$\log \big\langle e^{uZ} \big\rangle = \sum_{n=0}^{\infty} \sum_{\substack{\text{Connected } M \\ n \text{ boundaries}}} \frac{u^n}{n!} e^{S_0 \tilde{\chi}(M)}. \tag{3.13}$$

Furthermore, since the factor $\frac{u^n}{n!}$ is determined entirely by n while the factor $e^{S_0 \tilde{\chi}(M)}$ depends only on the genus g, the double sum in (3.13) may be written as the product

$$\log \big\langle e^{uZ} \big\rangle = \left(\sum_{g=0}^{\infty} e^{S_0 \chi(\tilde{M})} \right) \left(\sum_{n=0}^{\infty} \frac{u^n}{n!} \right) = \lambda e^u. \tag{3.14}$$

Here the last equality has used (3.9) to identify λ with the sum over g. We can extract the correlators $\Big\langle Z^n \Big\rangle$ by expanding the generating function $\exp(\lambda e^u)$ in powers of u.

We pause to note that there is a more direct way to compute the amplitudes $\langle Z^n \rangle$. Here we first divide by $\mathfrak{Z}$ to remove contributions from closed manifolds and thus any mention of $\mu(M)$. What remains is then just to simply count the relevant configurations remaining in (3.5). Such configurations are classified according to which of the n boundaries lie in the same connected component of spacetime, and thus by a partition of the set $\{1, 2, \ldots, n\}$ labelling the boundaries. For each connected component of spacetime, it then remains only to sum over genus, giving a factor of λ from (3.9). We may thus compute the amplitudes from a counting of partitions, graded by the number of subsets of $\{1, 2, \ldots, n\}$ that the partition defines:

$$\mathfrak{Z}^{-1}\langle Z^n \rangle = \sum_{\substack{\text{Partitions } p \\ \text{of } \{1,2,\ldots,n\}}} \lambda^{(\text{Number of subsets in } p)} = B_n(\lambda). \tag{3.15}$$

Here B_n is known as the Bell polynomial of order n (`BellB[n,λ]` in Mathematica; also called Touchard polynomial). In agreement with our previous result, these polynomials are indeed known to have the generating function $\exp(\lambda(e^u - 1))$ as in (3.14) after dividing by $\mathfrak{Z} = e^\lambda$.

To illustrate the counting in detail, consider the example of the third moment $\langle Z^n \rangle$; i.e., the case $n = 3$. There are five distinct ways to divide the three boundaries into connected components:

$\mathfrak{Z}^{-1}\langle Z^3 \rangle =$  (3.16)

$= \lambda^3 + 3\lambda^2 + \lambda$

Since the boundaries are distinguishable, the three configurations with two connected components are counted separately, and there are no explicit symmetry factors in the first line above.[10] The alternative counting used in (3.7) would instead list each topologically distinct term in (3.16) only once, but would accompany each term by the number N_L of distinct ordered lists that one can construct from the connected components and the factor of $\frac{n!}{m! \prod n_i!}$ from (3.7). This gives the identical result

$\mathfrak{Z}^{-1}\langle Z^3 \rangle = \frac{3!}{3!(1!)^3}$  $+ \frac{2\ 3!}{2!2!1!}$  $\frac{3!}{1!3!}$  (3.17)

$= \lambda^3 + 3\lambda^2 + \lambda,$

where the first term has $(N_L, m!, \frac{n!}{\prod n_i!}) = (1, 3!, \frac{3!}{1!1!1!})$ since the 3 components are all identical but have only one boundary each, the second term has $(N_L, m!, \frac{n!}{\prod n_i!}) =$

[10] For indistinguishable boundaries the answer would be multiplied by $\frac{1}{3!}$, or more generally by $\frac{1}{n!}$ for n boundaries).

$(2, 2!, \frac{3!}{2!1!})$ since the two components are not homeomorphic but the cylinder has 2 boundaries, and the third term has $(N_L, m!, \frac{n!}{\prod n_i!}) = (1, 1, \frac{3!}{3!})$ since all 3 boundaries lie in the single connected component.

We now interpret the amplitudes in terms of a probability distribution where Z is regarded as a random variable. To do this, we divide the generating function $\langle e^{uZ} \rangle$ by the normalisation factor $\mathfrak{Z}$ and write the result as the Taylor series for the exponential:

$$\mathfrak{Z}^{-1} \langle e^{uZ} \rangle = \sum_{d=0}^{\infty} p_d(\lambda) e^{ud}, \qquad p_d(\lambda) = e^{-\lambda} \frac{\lambda^d}{d!} . \tag{3.18}$$

Extracting the coefficient of $\frac{u^n}{n!}$ from (3.18) gives

$$\mathfrak{Z}^{-1} \langle Z^n \rangle = \sum_{d=0}^{\infty} d^n p_d(\lambda), \qquad p_d(\lambda) = e^{-\lambda} \frac{\lambda^d}{d!} , \tag{3.19}$$

showing that all moments can be generated from a single distribution for Z with support on nonnegative integers d having manifestly non-negative probabilities $\Pr(Z = d) = p_d(\lambda)$. We thus identify Z as a Poisson random variable with mean λ. We may also read this off directly from (3.14) using the fact that $\exp[\lambda(e^u - 1)]$ is the moment generating function for a Poisson random variable. Alternatively, one can see this from the amplitudes (3.15) using the fact that B_n is the nth moment of the Poisson distribution. The appearance of the Poisson distribution can be understood from the result that all connected components of spacetime contribute the same amplitude λ after summing over genus, independent of the number of boundaries. This corresponds to the fact that the cumulants of the Poisson distribution (that is, the completely connected correlation functions) are all equal to λ.

This is a surprising and remarkable result. As reviewed in section 5 below, a perturbative description of the theory following [57] (based on a Fock space labelled by number of baby universes and with wormholes treated as a small correction) would have led to the expectation that Z should have a continuous distribution supported on all real numbers. Instead, from our exact nonperturbative solution we find that the support of Z is discrete, and limited to nonnegative values.

Furthermore, for our choice $S_\partial = S_0$ (or more generally for $S_\partial = S_0 + \log n$ for any positive integer n), since Z takes nonnegative integer values d we find that the result is compatible with the interpretation (3.8) in terms of an ensemble of dual Hilbert spaces. Although at this stage this result appears to depend on fine tuning the parameter S_∂, we will see in section 3.7 that full consistency (in particular full reflection positivity) of the model in fact favours precisely the relation $S_\partial = S_0 + \log n$.

As a final comment, it is interesting that the relation (3.9) between the 'bare' parameter e^{S_0} and the physically observable parameter λ is not injective, but is instead

two-to-one. This means that there for a given value of e^{S_0}, there is a second value $e^{\tilde{S}_0}$ that gives rise to the same λ, and hence the same theory. In particular, we find

$$e^{-\tilde{S}_0} = 1 - e^{-S_0} \,. \tag{3.20}$$

This is a strong–weak self-duality of the model in the sense that the semiclassical limit of large S_0 suppresses connected topologies (and thus describes weakly coupled universes), but yields the same theory as a very small value of the dual $\tilde{S}_0$. At the self-dual value $e^{S_0} = 2$ we have $\lambda = 4$, and smaller values of λ correspond to complex couplings, with $e^{-S_0} \in \frac{1}{2} + i\mathbb{R}$. From the point of view of the path integral in a semiclassical expansion it is surprising that such a complex coupling gives rise to reflection positive amplitudes, and hence to a unitary Hilbert space and positive probabilities.

3.3 The baby universe Hilbert space

We can now give a complete description of the Hilbert space of closed universes $\mathcal{H}_{\rm BU}$. Every state can be written as a linear combination of $\left|Z^m\right\rangle$ created by inserting m boundaries in the past, with inner product

$$\begin{aligned} \left\langle Z^n \middle| Z^m \right\rangle &= \left\langle Z^{m+n} \right\rangle \\ &= e^{\lambda} B_{m+n}(\lambda) \\ &= \sum_{d=0}^{\infty} \frac{\lambda^d}{d!} d^{m+n}. \end{aligned} \tag{3.21}$$

A more general state $\sum_{n=0}^{\infty} c_n |Z^n\rangle$ can then be represented as $|f(Z)\rangle$, where f is a function with Taylor coefficients c_n, which grow slowly enough for convergence. Demanding that the partial sums $\left\{\sum_{n=0}^{N} c_n |Z^n\rangle\right\}_N$ form a Cauchy sequence guarantees that f defines an entire analytic function (see appendix A.1). Before considering the details of the inner product, we are thus led to the idea that $\mathcal{H}_{\rm BU}$ is a space of functions $f : \mathbb{R} \to \mathbb{C}$ (or perhaps $f : \mathbb{C} \to \mathbb{C}$), with argument Z.

We can read off the extension of the inner product to states $|f(Z)\rangle$ from the last line in (3.21):

$$\left\langle g(Z) \middle| f(Z) \right\rangle = \sum_{d=0}^{\infty} \frac{\lambda^d}{d!} \overline{g(d)} f(d). \tag{3.22}$$

This is (up to normalisation factor e^{λ}) the covariance of random variables $f(Z), g(Z)$ where Z is Poisson distributed. But the salient feature of (3.22) is that it depends only on the vales of f and g evaluated at non-negative integers (also known as the set $\mathbb{N}$ of

natural numbers). In particular, we find that the state $\big|f(Z)\big\rangle$ has zero norm whenever the function f vanishes on $\mathbb{N}$:

$$\Big\| \big|f(Z)\big\rangle \Big\|^2 = 0 \iff f(d) = 0 \text{ for all } d \in \mathbb{N}. \tag{3.23}$$

To form the Hilbert space $\mathcal{H}_{\rm BU}$, we must quotient by such null states as in (2.9). For example, since $\sin(\pi Z)$ vanishes on $\mathbb{N}$ we have the otherwise surprising relation

$$|\sin(\pi Z)\rangle = \sum_{n=0}^{\infty} \frac{(-1)^n \pi^{2n+1}}{(2n+1)!} \big|Z^{2n+1}\big\rangle = 0. \tag{3.24}$$

More generally, for any f we have $\big|\sin(\pi Z) f(Z)\big\rangle = 0$, so in some sense the space of null states is the same size as the total space before the quotient. Similarly, the Hartle-Hawking state can be represented by the constant function $f(Z) = 1$, or more generally by any function that has $f(d) = 1$ for all $d \in \mathbb{N}$ (for example, $|\mathrm{HH}\rangle = |e^{2\pi i j Z}\rangle$ for any integer j). To emphasise the impact of the quotient by null states, note that by adding vectors of the form $|Z^n \sin(\pi Z)\rangle$ we can change any finite number of coefficients c_n (for $n \neq 0$) in the expansion of the state $\sum_{n=0}^{\infty} c_n |Z^n\rangle$ at will. As a result, the only physical information in any finite collection of coefficients c_n is the overlap with the $Z = 0$ eigenstate (given by c_0).

These considerations reveal an enormous degeneracy in how states of $\mathcal{H}_{\rm BU}$ are represented as sums of $|Z^n\rangle$. We regard this degeneracy as a gauge equivalence. As described in section 2.2 this gauge symmetry is a natural modification of diffeomorphism invariance associated with allowing topology change in the functional integral. But the enormous power of this seemingly natural modification comes as a surprise. This indicates that the corrections to diffeomorphism invariance are not generic, but are instead highly correlated. As a result, the corrections conspire to enhance the impact of the gauge symmetry, and thus to produce the degeneracy observed above. Such conspiracies call out for a more fundamental explantation, and we will see in sections 3.7 and 4 below that at least some of these conspiracies are in fact implied by reflection positivity of our path integral.

In parallel with the treatment in section 2.3, we can now discuss the α-states of our model. These are the eigenstates $\big|Z = d\big\rangle$ of $\widehat{Z}$, labelled by $d \in \mathbb{N}$, and they must form a basis for $\mathcal{H}_{\rm BU}$. When expressed as a sum of the states $|Z^n\rangle$ states, we may choose coefficients defining the Taylor series of any analytic function taking a non-zero value at $Z = d$ but vanishing at other natural numbers, since multiplication by Z acts as multiplication by the constant d on such a function. One of the infinitely many ways

to represent such eigenstates states is then

$$\Big|Z = d\Big\rangle = \left(\frac{\lambda^d}{d!}\right)^{-1/2} \left|\frac{\sin(\pi Z)}{\pi(Z-d)}\right\rangle , \tag{3.25}$$

where the coefficient is chosen to enforce the normalisation

$$\Big\langle Z = d'\Big|Z = d\Big\rangle = \delta_{dd'} \ . \tag{3.26}$$

Finally, we discuss the spacetime interpretation of our operator $\widehat{Z}$ and its eigenstates $\big|Z = d\big\rangle$. From (3.22), note that projecting the states $\big|f(Z)\big\rangle$ onto the (here, one-dimensional) subspace where $\widehat{Z}$ takes the value d is equivalent to restricting the sum on the right-hand side of (3.22) to the given eigenvalue d, or equivalently to terms of order λ^d. But due to (3.9) (and the fact that the analogous equations are identical for any fixed number $n > 0$ of boundaries on the connected surface), these give precisely the contributions in (3.5) that arise from spacetimes with d connected components. We thus find that working in the eigenspace with eigenvalue d is equivalent to restricting the sum over amplitudes to terms where the universe has precisely d connected components[11].

In other words, the operator $\widehat{Z}$ counts the number of connected components of spacetime! This is quite surprising, since this is not a quantity we would naturally associate with a Cauchy slice if we were to attempt to quantise by gauge fixing diffeomorphisms (unlike the number of connected components of *space*, which is a natural observable when universes cannot split and join, but is not gauge invariant when they can).

The α-states are designed to make amplitudes factorise (2.23), and it is interesting to note how our model achieves this. To work in an α-state $\big|Z = d\big\rangle$, we can impose the nonlocal constraint that spacetime has exactly d connected components. This does not exclude wormhole configurations connecting multiple boundaries, but provides additional correlations between disconnected configurations of boundaries. It thus achieves factorisation in a surprising way, which may be instructive for less simple models. Note that our choice of symmetry factors on spacetimes without boundary, which otherwise only acts to renormalise $\mathfrak{Z}$, is crucial for this simple description of α-state correlation functions.

Since $\widehat{Z}$ takes values in $\mathbb{N}$, $\mathcal{H}_{\rm BU}$ has a natural representation as a harmonic oscillator Hilbert space in which $\widehat{Z}$ acts as a number operator.[12] We can define the annihilation

[11] We thank Xi Dong for discussions on this point.

[12] This is not to be confused with the free Fock space description of section 5, in which $\widehat{Z}$ is a harmonic oscillator position operator.

operator a as acting to shift functions of Z,

$$a\big|f(Z)\big\rangle = \sqrt{\lambda}\big|f(Z+1)\big\rangle, \tag{3.27}$$

so that we have the relations

$$\begin{aligned} \widehat{Z} &= N = a^\dagger a\,, \\ a|Z=0\rangle &= 0\,, \quad \text{and} \\ |Z=d\rangle &= \frac{1}{\sqrt{d!}}(a^\dagger)^d|Z=0\rangle. \end{aligned} \tag{3.28}$$

In this description, the Hartle-Hawking state is a coherent state, which can be represented as

$$\big|\,\mathrm{HH}\,\big\rangle = e^{\sqrt{\lambda}a^\dagger}|Z=0\rangle. \tag{3.29}$$

The distribution of the associated ensemble then follows from the well-known fact that the number operator follows a Poisson distribution in a coherent state.

3.4 End-of-the-world branes

We now extend the model described above by introducing dynamical boundaries, which (following [49]) we call end-of-the-world (EOW) branes. We choose to include an arbitrary number k of species of EOW brane, so each of these boundaries is labelled by an index $i \in \{1, 2, \ldots, k\}$. Equivalently, we can place a topological quantum mechanics on the EOW branes, with zero Hamiltonian and a k-dimensional Hilbert space, so that i labels an orthonormal basis of states in that Hilbert space. Apart from the species label, the only local data on an EOW brane is an orientation compatible with the spacetime it bounds.

Introducing the EOW branes has two effects. Firstly, they can appear as closed boundaries in the sum over topologies, but this is largely unimportant, only acting to change the value of λ so that it is no longer given by (3.9). More importantly, the EOW branes allow us to impose a new class of possible boundary conditions. Namely, we can specify that we have a boundary condition which is an oriented interval labelled at its endpoints by EOW brane species i and j. Since the interval is oriented, we may refer to it as having a past endpoint that creates an EOW brane of type i and a future endpoint that destroys an EOW brane of type j. We refer to both past and future labels as EOW brane sources. In a putative 0+1 dual, the condition that a boundary creates an EOW brane with label i corresponds to the preparation of a certain 0+1 dual state ψ_i. We denote a boundary interval between EOW branes i and j by (ψ_j, ψ_i) since the bulk path integral with this boundary condition should compute the inner

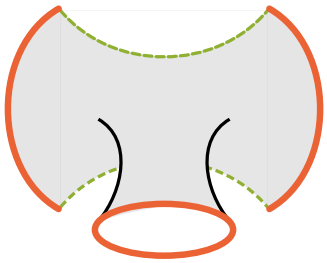
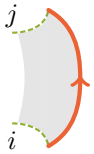


Figure 4: A spacetime contributing to an amplitude $\big\langle(\psi_j,\psi_i)(\psi_i,\psi_j)Z\big\rangle$. The solid red lines indicate asymptotically AdS boundaries, and the dashed green lines are EOW brane boundaries. The spacetime has two boundary components, each with the topology of a circle. One (solid red circle at bottom) is a single circular asymptotically AdS boundary (a Z-boundary). The other is formed by a pair of asymptotically AdS segments connected by a pair of EOW brane segments to form a topological circle.

product between these states.

$$(\psi_j,\psi_i) = \tag{3.30}$$



Since the boundaries carry an orientation, the notation distinguishes bra-vectors from ket-vectors so that $(\psi_j,\psi_i)\neq(\psi_i,\psi_j)$; in general, these are CPT conjugate boundary conditions. This coincides with the general notation introduced in section 2.2.

Including the ψ_i, the most general amplitude can now be written

$$\Big\langle Z^m(\psi_{j_1},\psi_{i_1})\cdots(\psi_{j_n},\psi_{i_n})\Big\rangle. \tag{3.31}$$

The associated boundary conditions for the path integral require m circular boundaries without EOW brane sources and n additional interval boundary segments labelled appropriately with EOW brane species. Since the EOW branes are dynamical, the path integral is then computed by summing over all oriented surfaces whose circular boundaries are of the following three types: 1) circular EOW brane boundaries, each labelled by an arbitrary species independent of all boundary conditions, 2) m circular boundaries without EOW brane labels as dictated by the number of Z's in the amplitude, and 3) additional circular boundaries formed by partitioning into subsets the oriented intervals (ψ_j,ψ_i) dictated by the boundary conditions and, for each subset, forming a circle by connecting the (ψ_j,ψ_i) segments using oriented EOW brane segments whose species labels match the source labels at both endpoints. See figure 4 for an example.

We now know the set of amplitudes to compute and the corresponding configurations over which we are to sum. It remains only to specify the measure on the configurations. As before, the Euler characteristic is the unique local action without introducing additional degrees of freedom. However, we will again include a parameter S_∂ associated with each circular boundary. We use the same S_∂ for every circular boundary, no matter how it is formed from asymptotic pieces and EOW branes. Again, we will see in section 3.7 that this can be obtained by introducing additional local degrees of freedom which reside on both asymptotic and EOW brane boundaries, and integrating them out. While this no longer corresponds to a simple scaling of our operators, we will nonetheless once again focus on the case $S_\partial = S_0$, resulting in an action which counts only genus and not the number of boundary components, and comment on the extension to other values in section 3.7.

It remains to specify the symmetry factors that will be the analog of $\mu(M)$ in (3.1). In doing so, it is useful to note that, since all asymptotic boundaries are treated as distinguishable, they will not contribute to symmetry factors. The only indistinguishable boundaries are those formed by circles involving EOW branes alone. Furthermore, such circles are completely independent of the boundary conditions. They thus enter all of our sums in precisely the same way as the genus g. The analogue of (3.5) for our new model is then

$$\left\langle Z^m(\psi_{j_1},\psi_{i_1})\cdots(\psi_{j_n},\psi_{i_n})\right\rangle = \sum_M \mu(M) e^{S_0\tilde{\chi}} \ , \tag{3.32}$$

where we sum over diffeomorphism classes of surface M with the boundary conditions specified on the left hand side. The measure μ is analogous to (3.2) but includes additional factors associated with counting end-of-the-world branes using Bose statistics.

We may now proceed to evaluate the above amplitudes. As a first step, we again define λ as the sum over connected surfaces with no asymptotic boundaries in analogy with (3.9). However, this sum must now allow for the possibility of circular EOW brane boundaries, each with k possible species labels. Since EOW brane boundaries can be specified in precisely the same way for each genus, this simply multiplies the result (3.9) by an overall factor counting the number of possible such labelled boundaries. For a fixed number n of EOW brane boundaries, including symmetry factors we count $\frac{k^n}{n!}$ ways to label the boundaries with k species. Summing this factor over n shows the new factor to be e^k and we obtain

$$\lambda = \frac{e^{2S_0}}{1-e^{-2S_0}} e^k. \tag{3.33}$$

As before, we can now compute all amplitudes through a generating function, where we sum over all configurations, with any number of asymptotic boundaries, and

fugacities u and t_{ij} (with $i = 1, \cdots, k$) for the Z and (ψ_j, ψ_i) boundaries respectively. As we explain below, this yields

$$\left\langle \exp\left(uZ + \sum_{i,j=1}^{k} t_{ij}(\psi_j, \psi_i)\right)\right\rangle = \exp\left[\lambda \frac{e^u}{\det(I-t)}\right], \tag{3.34}$$

where t is the $k \times k$ matrix with entries t_{ij}, and I the $k \times k$ identity matrix.

Once again, we compute this result by writing it as the exponential of a sum over connected spacetimes, each weighted by a factor of λ from summing over genus and closed EOW branes. The connected contribution is a sum over all possible boundaries we could insert on a given connected spacetime (excepting circular EOW brane boundaries, which have already been absorbed into λ). This sum is itself given as the exponential of a sum over distinct types of boundaries:

$$\frac{e^u}{\det(I-t)} = \exp\left[u + \sum_{n=1}^{\infty} \frac{1}{n} \operatorname{Tr} t^n\right] \tag{3.35}$$

The u accounts for insertions of circle boundaries Z as before. The nth term in the sum comes from boundary components consisting of n intervals corresponding to some (ψ_j, ψ_i), alternating with n EOW branes. Summing over species of EOW branes results in the matrix product and trace, and the factor of $\frac{1}{n}$ avoids overcounting equivalent configurations where the n component intervals are cyclically permuted.

For an alternative route to this result where various factors are more explicit, we can present (3.32) as a sum over ordered lists of connected manifolds. This is readily obtained from (3.7) by recognizing that the circular EOW brane boundaries enter every sum on the same footing with the genus g. We have

$$\begin{aligned}&\left\langle Z^m(\psi_{j_1}, \psi_{i_1}) \cdots (\psi_{j_n}, \psi_{i_n})\right\rangle \\ &\quad = \sum_{\substack{L, \{D_i, I_i\} \\ 1 \le i \le L \text{ with } \sum_i I_i = n_I \\ \sum_i D_i = D}} \sum_{\substack{\text{Ordered lists } M_L \\ \text{of } L \text{ connected surfaces} \\ \text{where entry } i \text{ has } D_i, I_i \\ \text{distinguishable/indistinguishable} \\ \text{boundaries}}} C(D) \frac{u^m}{L!} \frac{k^{\sum_i I_i}}{\prod_i I_i!} \frac{D!}{\prod_i D_i!} e^{S_0 \tilde{\chi}},\end{aligned} \tag{3.36}$$

where the factor $\frac{k^{I_i}}{I_i!}$ for each connected manifold counts the number of ways (including symmetry factors) to assign EOW brane labels to I_i indistinguishable circular boundaries and the factor $\frac{D!}{\prod_i D_i!}$ again counts partitions of the D distinguishable boundaries into (labelled) subsets of size D_i. Finally, the factor $C(D)$ represents the number of ways to form D distinguishable boundaries from the specified boundary conditions (together with interpolating EOW brane segments).

In comparing with (3.36), the relation to the exponential of (3.34) is clear from the factor of $1/L!$ in (3.36), the inclusion of factors of $\frac{k_i^I}{I_i!}$ in (3.34), and the defining property of generating functions. By this last feature, we mean the fact that the definition of the generating functions (3.34) converts the factors $C(D)\frac{D!}{\prod_i D_i!}$ counting the number of ways to match distinguishable boundaries to boundary conditions into the above-described weighted sum over all possible boundary conditions for each connected component.

We now interpret the amplitudes as describing an ensemble, for which (3.34) is the (unnormalised) generating function for moments of random variables Z and (ψ_j, ψ_i). Let us first set $t = 0$ in order to consider the marginal distribution of Z. We then recover the old result (3.14) without EOW branes, so Z is again Poisson distributed, though with a new value of λ given by (3.33).

We can now characterise the distribution of (ψ_j, ψ_i) by conditioning on $Z = d$ for each fixed $d \in \mathbb{N}$. To find the corresponding conditional generating functions, we Taylor expand the exponential in (3.34) and write each term as an average over the Poisson probabilities $p_d(\lambda) = e^{-\lambda}\frac{\lambda^d}{d!}$:

$$\begin{aligned}\left\langle \exp\left(uZ + \sum_{i,j=1}^{k} t_{ij}(\psi_j, \psi_i)\right)\right\rangle &= e^{\lambda}\sum_{d=0}^{\infty} e^{ud} p_d(\lambda)\left\langle \exp\sum_{i,j=1}^{k} t_{ij}(\psi_j, \psi_i)\right\rangle_{Z=d} \\ \implies \left\langle \exp\left(\sum_{i,j=1}^{k} t_{ij}(\psi_j, \psi_i)\right)\right\rangle_{Z=d} &= \det(I-t)^{-d}. \end{aligned} \tag{3.37}$$

The result is the generating function for a standard complex Wishart distribution [91] with d degrees of freedom.

To make this more transparent, and to simultaneously explain this distribution to the uninitiated reader, we can rewrite the generating function by introducing kd 'auxiliary' complex variables ψ_i^a, arranged in a $d \times k$ matrix. The index $i = 1, \ldots, k$ labels the EOW brane states, and we will interpret $a = 1, \ldots, d$ as labels for an orthonormal basis of the boundary Hilbert space $\mathcal{H}_{\text{CFT}}$ (which is d-dimensional based on our interpretation (3.8) of Z). The ψ_i^a variables will be interpreted as the components of the EOW brane states ψ_i in this orthonormal basis.

In terms of the ψ_i^a variables, our Wishart generating function (3.37) can now be written as a Gaussian integral:

$$\det(1-t)^{-d} = \int \prod_{i=1}^{k}\prod_{a=1}^{d}\left(\frac{1}{\pi} d\psi_i^a d\bar{\psi}_i^a\, e^{-\bar{\psi}_i^a \psi_i^a}\right) \exp\left(\sum_{i,j=1}^{k} t_{ij}\sum_{a=1}^{d}\bar{\psi}_j^a\psi_i^a\right) \tag{3.38}$$

Comparing with the expectation value (3.37) we are computing, we can read off the

distribution by identifying the matrix of inner products (ψ_j, ψ_i) as

$$(\psi_j, \psi_i) = \sum_{a=1}^{d} \bar{\psi}_j^a \psi_i^a \tag{3.39}$$

from the final factor in the integral. The remainder of the integral gives the measure for the ψ_i^a, as independent random variables, each chosen from a complex normal (Gaussian) distribution with unit variance:

$$\psi_i^a \sim \text{ independent standard complex normal random variables.} \tag{3.40}$$

In the 0+1 dual interpretation, this means that the wavefunction of each EOW brane states is selected independently and uniformly at random from the unit sphere of a d-dimensional Hilbert space $\mathcal{H}_{\text{CFT}}$, and then multiplied by a random normalization so that its squared norm is drawn from an appropriate χ^2-distribution. In particular, the number of linearly independent states, given by the rank of the matrix of inner products, is bounded by Z: with probability one we have

$$\text{rank}(\psi_j, \psi_i) = \min\{k, Z\}. \tag{3.41}$$

This is another surprising and remarkable result from such a simple model, since in the semiclassical limit (without the exponentially small effects of spacetime wormholes) the k EOW brane states appear to be orthogonal, and we can choose k to be as large as we like. As discussed below in section 5, even if we include Euclidean wormholes there is an expansion in e^{-S_0} which for a finite number of amplitudes at any finite order gives no obvious sign that apparently distinct EOW brane states must in fact be linearly dependent. Nonetheless, in the complete solution after summing all nonperturbative effects, we find that the number of linearly independent states is truncated. As in [49], as and discussed further in section 4, this is a version of the semiclassical Page curve [1].

At first sight, this appears to require an enormous conspiracy in the nonperturbative contributions, which might lead one to suspect that it is an artefact of studying particularly simple models. We will show below that this is not the case, since it follows from a more primitive principal, namely reflection positivity of the path integral. For this, we must study the Hilbert space interpretation of the model with EOW branes.

3.5 Baby universe Hilbert space with EOW branes

We now incorporate the EOW branes into the baby universe Hilbert space. This enlarges the space relative to that of section 3.3 because, along with circular closed

universes, we also have k^2 new types of universe whose spatial slice is an interval bounded by EOW branes, say with labels i and j (where the orientation defines a preferred order). On the other hand, the above-mentioned conspiracies will also imply the existence of new null states.

It is most straightforward to construct $\mathcal{H}_{\text{BU}}$ from the α-states. These are eigenstates of the $\widehat{Z}$ operator as before, but now are simultaneously eigenstates of the k^2 operators $\widehat{(\psi_j, \psi_i)}$ as well; note that Hermitian conjugation acts on these operators by swapping i, j. We label the corresponding eigenvalues by Z_α and $(\psi_j, \psi_i)_\alpha$, so we have

$$\begin{aligned} \widehat{Z}\big|\alpha\big\rangle &= Z_\alpha \big|\alpha\big\rangle \\ \widehat{(\psi_j, \psi_i)}\big|\alpha\big\rangle &= (\psi_j, \psi_i)_\alpha \big|\alpha\big\rangle. \end{aligned} \tag{3.42}$$

The set of α-states is determined by the allowed sets of eigenvalues, which is constrained by (3.41).

As in section 3.3, the eigenvalues Z_α of $\widehat{Z}$ are given by the nonnegative integers d. Indeed, we can still define states $|Z = d\rangle$ by any of the means discussed in that section, for example by (3.25). However, they are now not full α-states, since they are eigenstates only of $\widehat{Z}$ and not of $\widehat{(\psi_j, \psi_i)}$. Instead they are the projections of the Hartle-Hawking state onto the corresponding eigenspace of $\widehat{Z}$. We can generate the rest of this eigenspace by acting with the operators $\widehat{(\psi_j, \psi_i)}$ on $|Z = d\rangle$.

In each such eigenspace, we can now diagonalise the operators $\widehat{(\psi_j, \psi_i)}$. Their simultaneous eigenvalues correspond to Hermitian $k \times k$ positive definite matrices of rank at most d (though any rank other than $\min(d, k)$ has probability zero in any normalizable state). The baby universe Hilbert space therefore decomposes as a direct sum:

$$\begin{aligned} \mathcal{H}_{\text{BU}} &= \bigoplus_{d=0}^{\infty} \mathcal{H}_{Z=d} \\ \mathcal{H}_{Z=d} &= L^2(M_k^d) \\ M_k^d &= \{\text{Hermitian p.d. } k \times k \text{ matrices}, \text{rank} \leq d\}. \end{aligned} \tag{3.43}$$

The summands $\mathcal{H}_{Z=d}$ are the usual L^2 spaces of square integrable functions on the relevant space of restricted rank matrices M_k^d (defined with any convenient smooth measure). For $d \leq k$, M_k^d forms a $(2kd - d^2)$-dimensional manifold; we can write $(\psi_j, \psi_i) = \sum_{a=1}^{d} \bar{\psi}_j^a \psi_i^a$ as in (3.39) so that the $2kd$ counts the number of independent real parameters in ψ_i^a while the d^2 subtracts for the invariance under unitary rotations of the a directions. For $d \geq k$, the restriction on rank is vacuous.

With this description, the α-states are delta function wavefunctions living in the subspaces $\mathcal{H}_{Z=d}$, supported on some particular matrix $(\psi_j, \psi_i)_\alpha \in M_k^d$. In particular,

we write their inner product as

$$\langle \alpha' | \alpha \rangle = \delta_{\alpha\alpha'}, \tag{3.44}$$

where $\delta_{\alpha\alpha'}$ is the product of a Kronecker delta $\delta_{Z_\alpha Z_{\alpha'}}$ for the eigenvalue of $\widehat{Z}$ with an appropriate Dirac delta function on M_k^d associated with the choice of L^2 measure in (3.43).

Finally, the wavefunction of the Hartle-Hawking state in this description is given by

$$\big\langle \alpha \big| \mathrm{HH} \big\rangle = \sqrt{\frac{\lambda^{Z_\alpha}}{Z_\alpha!} f_{Z_\alpha}\big((\psi_j, \psi_i)_\alpha\big)} \ , \tag{3.45}$$

where f_{Z_α} is the probability density function of the complex Wishart distribution with Z_α degrees of freedom with respect to the measure on our L^2 space; this is the overlap $\big\langle \alpha \big| Z = Z_\alpha \big\rangle = f_{Z_\alpha}$. For $Z_\alpha \geq k$, this density is given explicitly in (3.65).

3.6 Hilbert spaces with boundaries

Our discussion of Hilbert spaces is not yet complete. In particular, other Hilbert spaces of interest arise when we insert complete sets of states on Cauchy slices that intersect 'asymptotically AdS' boundaries. Here there are two types of boundary, distinguished by their orientation; we call them 'left' and 'right' boundaries of space. In a 0+1 dual, the two types of boundaries would correspond to CPT conjugate theories.

In our model, the most general slice Σ of the asymptotically AdS boundaries will consist of n_L left boundaries and n_R right boundaries. We thus denote the associated Hilbert space $\mathcal{H}_\Sigma$ from section 2.4 as $\mathcal{H}_{n_L,n_R}$. Reversing the orientation of all boundaries gives the dual (Hermitian conjugate) Hilbert space, so $\mathcal{H}^*_{n_L,n_R} = \mathcal{H}_{n_R,n_L}$. The simplest of these is $\mathcal{H}_{\mathrm{BU}} = \mathcal{H}_{0,0}$, which we have already discussed. We will be primarily interested in the one-sided Hilbert space $\mathcal{H}_{0,1}$ (related to $\mathcal{H}_{1,0}$ by duality) and the two-sided space $\mathcal{H}_{1,1}$.

We begin by considering the single boundary Hilbert space $\mathcal{H}_{0,1}$, which is spanned by states of the form $|\psi_i; Z^m\,(\psi_{j_1}, \psi_{i_1}) \cdots (\psi_{j_n}, \psi_{i_n})\rangle$. Recall that the operator $\widehat{\psi}_i$ maps $\mathcal{H}_{\mathrm{BU}}$ to $\mathcal{H}_{0,1}$ (or more generally $\mathcal{H}_{n_L,n_R} \to \mathcal{H}_{n_L,n_R+1}$). All of the above states can be produced by acting with the operator $\widehat{\psi}_i$ on a state of closed baby universes. In particular, we can span $\mathcal{H}_{0,1}$ by acting with one of the k operators $\widehat{\psi}_i$ (for $i = 1, \ldots k$) on α-states of $\mathcal{H}_{\mathrm{BU}}$. The inner product on such states is

$$\big\langle \psi_j; \alpha' \big| \psi_i; \alpha \big\rangle = \big\langle \alpha' \big| \widehat{(\psi_j, \psi_i)} \big| \alpha \big\rangle = \delta_{\alpha\alpha'} (\psi_j, \psi_i)_\alpha \,, \tag{3.46}$$

so in particular, the different α-sectors are orthogonal, and $\mathcal{H}_{0,1}$ admits a direct sum decomposition

$$\mathcal{H}_{0,1} = \bigoplus_\alpha \mathcal{H}^\alpha_{0,1}, \tag{3.47}$$

(where this is to be understood in the appropriate sense given that some of the parameters defining α are continuous). The inner product on each sector $\mathcal{H}^\alpha_{0,1}$ is simply given by the matrix of eigenvalues $(\psi_j, \psi_i)_\alpha$. On sectors with $Z_\alpha < k$, this is degenerate, and $\mathcal{H}^\alpha_{0,1}$ is Z_α-dimensional:

$$\dim \mathcal{H}^\alpha_{0,1} = \min\{k, Z_\alpha\}. \tag{3.48}$$

Next, we look at the two-boundary sector $\mathcal{H}_{1,1}$. In the same way, this Hilbert space can be populated by acting with boundary creating operators on states of $\mathcal{H}_{\rm BU}$, for example on α-states. We have the same direct sum structure as before, $\mathcal{H}_{1,1} = \bigoplus_\alpha \mathcal{H}^\alpha_{1,1}$. States within each $\mathcal{H}^\alpha_{1,1}$ can be created by acting with separate EOW brane states on left and right boundaries using $\widehat{\psi^*_j}\widehat{\psi_i}$. But we now have an additional possibility where we introduce a single asymptotic boundary that connects left and right. In a general theory, one might call this the cylinder boundary (with topology Σ times an interval), and one might think of it as obtained by cutting in half a partition function on $\Sigma \times S^1$. By acting on $\big|\mathrm{HH}\big\rangle$, it thus creates a state that one expects to interpret as a 'thermofield double' in some CFT dual. In our case the cylinder degenerates to a line segment (since Σ is a point), which we can think of as half of a Z circle. We denote the boundary condition by $\smile$, the associated operator by $\widehat{\smile}$, and the resulting state by $\Big|\smile\Big\rangle = \widehat{\smile}\Big|\mathrm{HH}\Big\rangle$. Thus,

$$\mathcal{H}^\alpha_{1,1} \text{ is spanned by } \big|\psi^*_j, \psi_i; \alpha\big\rangle,\ \big|\smile; \alpha\big\rangle, \tag{3.49}$$

and the inner products of these states are given by

$$\begin{aligned}
\big\langle \psi^*_{j_2}, \psi_{i_2}; \alpha' \big| \psi^*_{j_1}, \psi_{i_1}; \alpha\big\rangle &= \delta_{\alpha\alpha'} (\psi_{i_2}, \psi_{i_1})_\alpha \, (\psi_{j_1}, \psi_{j_2})_\alpha, \\
\big\langle \smile; \alpha' \big| \psi^*_{j_1}, \psi_{i_1}; \alpha\big\rangle &= \delta_{\alpha\alpha'} (\psi_{j_1}, \psi_{i_1})_\alpha, \\
\big\langle \smile; \alpha' \big| \smile; \alpha\big\rangle &= \delta_{\alpha\alpha'} Z_\alpha.
\end{aligned} \tag{3.50}$$

From the first of these, we see that for fixed α the states $|\psi^*_j, \psi_i; \alpha\rangle$ span a subspace isomorphic to the tensor product of two single boundary subspaces, so this tensor product embeds naturally in $\mathcal{H}^\alpha_{1,1}$; i.e., $\mathcal{H}^\alpha_{0,1} \otimes \mathcal{H}^\alpha_{1,0} \subseteq \mathcal{H}^\alpha_{1,1}$. This inclusion could be an exact equality, but only if the new state $|\smile; \alpha\rangle$ can be built from a linear combination of factorised states $|\psi^*_j, \psi_i; \alpha\rangle$. This suggests that we look for a linear combination

$$\big|\Delta\big\rangle = \big|\smile; \alpha\big\rangle - \sum_{i,j=1}^{k} c_{ij} \big|\psi^*_j, \psi_i; \alpha\big\rangle \tag{3.51}$$

with zero norm. Such a vector would be projected out of the Hilbert space $\mathcal{H}_{1,1}$, giving $|\Delta\rangle = 0$ and providing an identity relating the cylinder state $\Big|\smile\Big\rangle$ to a superposition of one-sided states.

Before computing the norm of our ansatz $|\Delta\rangle$, we first change to a more convenient basis diagonalising the EOW brane inner product in the α-state in question (with eigenvalues $(\psi_j, \psi_i)_\alpha$). Specifically, we pick linear combinations ϕ_a of the ψ_i boundary conditions for which $(\phi_b, \phi_a)_\alpha = \delta_{ab}$, with the index $a = 1, \ldots, r$ running up to the rank of the matrix of inner products. In this basis, we rewrite our candidate null state and compute its norm:

$$\big|\Delta\big\rangle = |\smile; \alpha\rangle - \sum_{a,b=1}^{r} c_{ab} |\phi_b^*, \phi_a; \alpha\rangle \tag{3.52}$$

$$\begin{aligned}
\big\langle\Delta\big|\Delta\big\rangle &= \big\langle\smile; \alpha\big|\smile; \alpha\big\rangle - \sum_{a,b=1}^{r} c_{ab} \big\langle\smile; \alpha\big|\phi_b^*, \phi_a; \alpha\big\rangle \\
&\quad - \sum_{a,b=1}^{r} c_{ab}^* \big\langle\phi_b^*, \phi_a; \alpha\big|\smile; \alpha\big\rangle + \sum_{a,b,a',b'=1}^{r} c_{ab} c_{a'b'}^* \big\langle\phi_{b'}^*, \phi_{a'}; \alpha\big|\phi_b^*, \phi_a; \alpha\big\rangle \\
&= Z_\alpha - 2 \sum_{a=1}^{r} \operatorname{Re} c_{aa} + \sum_{a,b=1}^{r} |c_{ab}|^2 \\
&= Z_\alpha - r \qquad\qquad (c_{ab} = \delta_{ab}).
\end{aligned}$$

In the last line we have chosen the coefficients $c_{ab} = \delta_{ab}$ to be δ_{ab}, as this minimizes $\big\langle\Delta\big|\Delta\big\rangle$.

The above calculation teaches us two things. Firstly, for the norm to be nonnegative we have an inequality which applies in all α states:

$$\text{Reflection positivity} \implies Z_\alpha \geq \operatorname{rank}(\psi_j, \psi_i)_\alpha. \tag{3.53}$$

This explains our empirical result (3.41) that the rank of the EOW brane inner product is bounded by Z_α, in terms of reflection positivity of the path integral. The same argument can be used in much more general models, and we repeat it with the inclusion of a conserved energy in section 4, where we also connect it with the Page curve [1].

Secondly, we find that if the inequality (3.53) is saturated, we have $|\Delta\rangle = 0$, and hence an identity

$$|\smile; \alpha\rangle = \sum_{a=1}^{r} |\phi_a^*, \phi_a; \alpha\rangle. \tag{3.54}$$

Since the 'factorized states' $|\psi_j^*, \psi_i; \alpha\rangle$ then span the two-sided Hilbert space $\mathcal{H}_{1,1}^\alpha$, we also find an equivalence between Hilbert spaces

$$\mathcal{H}_{0,1}^\alpha \otimes \mathcal{H}_{1,0}^\alpha \equiv \mathcal{H}_{1,1}^\alpha \,. \tag{3.55}$$

This factorization holds in our model for sectors with $Z_\alpha \leq k$; i.e., when there are enough EOW branes to populate a one-sided Hilbert space of dimension Z_α.

To emphasise the importance of α-states in this argument, we examine how it fails in a more general (normalised) state $|\Psi\rangle \in \mathcal{H}_{\rm BU}$, such as the Hartle-Hawking state. Specifically, let us choose linear combinations ϕ_a of EOW brane states ψ_i to diagonalise the expectation value of the inner product in the state $|\Psi\rangle$; i.e., we take

$$\big\langle \Psi \big| \widehat{(\phi_b, \phi_a)} \big| \Psi \big\rangle = \delta_{ab} \,, \tag{3.56}$$

where $a, b = 1, \cdots, r$, with $r = \text{rank} \big\langle \Psi \big| \widehat{(\psi_j, \psi_i)} \big| \Psi \big\rangle$. If we now compute the norm of the state

$$\big| \Delta \big\rangle = \big| \circlearrowright ; \Psi \big\rangle - \sum_{a=1}^{r} \big| \phi_a^*, \phi_a; \Psi \big\rangle, \tag{3.57}$$

we find an extra term, coming from the overlaps $\big\langle \phi_b^*, \phi_b; \Psi \big| \phi_a^*, \phi_a; \Psi \big\rangle$:

$$\big\langle \Delta \big| \Delta \big\rangle = \big\langle \Psi \big| \widehat{Z} \big| \Psi \big\rangle - r + \sum_{a,b=1}^{r} \text{Var}_\Psi \left[(\phi_b, \phi_a) \right] . \tag{3.58}$$

Here we have defined the variance of boundary condition X as the connected amplitude for $XX^\dagger$,

$$\text{Var}_\Psi[X] = \big\langle \Psi \big| \widehat{X} \widehat{X}^\dagger \big| \Psi \big\rangle - \big\langle \Psi \big| \widehat{X} \big| \Psi \big\rangle \big\langle \Psi \big| \widehat{X}^\dagger \big| \Psi \big\rangle. \tag{3.59}$$

This vanishes in α-states, though is generically non-zero.

For example, in the Hartle-Hawking state, the expectation value of the overlaps of EOW brane states is already diagonal,

$$\frac{\Big\langle \text{HH} \Big| \widehat{(\psi_j, \psi_i)} \Big| \text{HH} \Big\rangle}{\Big\langle \text{HH} \Big| \text{HH} \Big\rangle} = \lambda \delta_{ij} \tag{3.60}$$

so we can define $\phi_a = \lambda^{-1/2} \psi_a$, and we have $r = k$. The variance of the individual terms (ϕ_b, ϕ_a) is small,

$$\text{Var}_{\rm HH}[(\phi_b, \phi_a)] = \lambda^{-1} (1 + \delta_{ab}), \tag{3.61}$$

but there are k^2 such terms, so they are collectively important when k is of order λ or larger. As a result, (3.58) gives no meaningful bound relating the rank of the inner product (ψ_j, ψ_i) to the partition function Z. Note that this is not really an issue of fluctuations in the particular parameter Z_α, as the same discussion applies to the states $|Z = d\rangle$, which fix the eigenvalue of $\widehat{Z}$ but not those of $\widehat{(\psi_j, \psi_i)}$.

Returning to the issue of reflection positivity, we should also discuss the Hilbert spaces $\mathcal{H}_{n_L, n_R}$ associated with arbitrary numbers of left and right boundaries. But

in our model all possible boundary conditions creating such states can be formed by combining $\smile$ with the above ψ_i. In superselection sectors with $Z_\alpha \leq k$, the above result then implies $\mathcal{H}_{n_L,n_R} = \mathcal{H}_{1,0}^{\otimes n_L} \otimes \mathcal{H}_{0,1}^{\otimes n_R}$ and the inner product on $\mathcal{H}_{n_L,n_R}$ is positive definite. In superselection sectors with $Z_\alpha > k$ the higher Hilbert spaces are not tensor products of the lower Hilbert spaces. But much as above, considering states similar to (3.57) again shows the inner product to be positive for $Z_\alpha > k = r$. We thus see by direct calculation that our path integral satisfies reflection positivity.

3.7 The boundary parameter S_∂

We now discuss the parameter S_∂, contributing an action proportional to the number of boundaries. First we describe how changing S_∂ from its preferred value $S_\partial = S_0$ alters the physics, and thus in particular explain why this value is preferred. We then discuss how we might naturally incorporate such a parameter in the model.

Let us first consider the model without EOW branes, discussed in sections 3.1, 3.2 and 3.3. There the only effect of S_∂ is to rescale the quantities and operators associated with the Z boundaries. We thus find an ensemble interpretation in which Z is $e^{S_\partial - S_0}$ times a Poisson random variable, so that the α-states are characterised by $\widehat{Z}$ eigenvalues $Z_\alpha \in e^{S_\partial - S_0}\mathbb{N}$. From the gravitational perspective, there is nothing wrong with this model for any positive value of S_∂. In particular, reflection positivity is preserved for all Hilbert spaces. Complex values are excluded by reflection positivity on $\mathcal{H}_{1,1}$, which is spanned by orthogonal states $|\smile;\alpha\rangle$ with norm $\langle\smile;\alpha|\smile;\alpha\rangle = Z_\alpha$. From the boundary perspective, there is a good dual interpretation only when $e^{S_\partial - S_0}$ is a nonnegative integer, so that Z_α takes nonnegative integer values which can be interpreted as the dimension of a dual Hilbert space. Nothing from the bulk perspective appears to prefer such values, so our choice $S_\partial = S_0$ appears to be rather artificial.

This changes once we introduce the EOW brane states. The bulk then provides a principled reason to prefer particular values of S_∂, as the inner product on EOW brane states will otherwise fail to be positive semidefinite. To see this, we focus on a sector of $\mathcal{H}_{\text{BU}}$ with fixed $\hat{Z}$ eigenvalue $Z = d$, in which our EOW brane amplitudes are given by the generating function (3.37), reproduced below with fugacities t_{ij} rescaled by a factor of i for later convenience and with the matrix of EOW brane inner products encoded in a $k \times k$ Hermitian matrix M, $M_{ij} = (\psi_j, \psi_i)$:

$$\chi_{d,k}(t) = \left\langle e^{i\,\mathrm{Tr}(tM)} \right\rangle_{Z=d} = \det(1 - it)^{-d} \tag{3.62}$$

For $d \in \mathbb{N}$, by introducing dk auxiliary Gaussian variables we showed in (3.38) that this gives a probability distribution for M, and hence a reflection positive inner product on $\mathcal{H}_{\text{BU}}$. This argument does not apply for $d \notin \mathbb{N}$, so we must find a different way to determine whether we have a positive semidefinite inner product.

If M is to be interpreted as a random variable selected from some probability distribution, (3.62) defines $\chi_{d,k}$ as the characteristic function of the distribution. This is the Fourier transform of the probability density function $p_{d,k}$, which is in general a distribution on the space of $k \times k$ Hermitian matrices. It thus determines our inner product, which acts on a space of functions f, g of $k \times k$ Hermitian matrices M:

$$\langle g|f\rangle = \int dM p_{d,k}(M) g(M)^* f(M), \tag{3.63}$$

$$\text{where } \chi_{d,k}(t) = \int dM e^{i \operatorname{Tr}(tM)} p_{d,k}(M). \tag{3.64}$$

The distribution $p_{d,k}$ is determined uniquely from the inverse Fourier transform of $\chi_{d,k}$.[13] For this to define a positive semidefinite inner product, we need $p_{d,k}$ to be a nonnegative distribution (that is, it gives positive values when integrated against positive test functions such as $|f(M)|^2$). The question of whether the $Z = d$ subspace of $\mathcal{H}_{\text{BU}}$ has a positive semidefinite inner product is equivalent to the existence of a probability distribution with characteristic function $\chi_{d,k}$.

A succinct summary answering this question is contained in [92], to which we refer the reader for the results we now use. For $d > k$, the inverse Fourier transform of $\chi_{d,k}$ is a continuous function of M, taking non-zero values only on positive-definite matrices:

$$\begin{aligned} p_{d,k}(M) &= \mathcal{N}_{d,k} \det(M)^{d-k} e^{-\operatorname{Tr} M}, \quad M \text{ positive definite}, \\ \mathcal{N}_{d,k}^{-1} &= \pi^{\frac{k(k-1)}{2}} \Gamma(d)\Gamma(d-1)\cdots\Gamma(d-(k-1)). \end{aligned} \tag{3.65}$$

This is manifestly nonnegative and so defines a probability distribution. This result extends to $d > k-1$, where the probability density diverges at the edge where M becomes degenerate, but is still integrable. This is easiest to see from the density in terms of the eigenvalues of M; fixing $k-1$ positive eigenvalues and taking the last $\lambda \to 0$, the density goes as λ^{d-k}. The important result for us is that this range $d > k-1$, along with the smaller nonnegative integer values of d already covered by (3.38), turns out to exhaust the values of d for which the inner product on $\mathcal{H}_{\text{BU}}$ is positive semidefinite:

$$\begin{aligned} \chi_{d,k}(t) = \det(1-it)^{-d} &\text{ defines a probability distribution} \\ &\iff d \in \{0, 1, 2, \ldots, k-2\} \cup [k-1, \infty). \end{aligned} \tag{3.66}$$

We can intuit this from (3.65) by analytic continuation of the density in d. As d approaches $k-1$, the density goes to zero for any fixed positive definite matrix from

[13] Our integration measure on Hermitian matrices is defined as the flat measure on independent real components, $dM = \prod_i dM_{ii} \prod_{i<j} d\operatorname{Re} M_{ij} d\operatorname{Im} M_{ij}$, and here we take t to be a Hermitian matrix so that $\operatorname{Tr}(tM)$ is real.

the zero in normalisation factor $\mathcal{N}_{d=k-1,k} = 0$, but the probability density piles up near $\det M = 0$ and we end up with a probability density supported on the submanifold of singular matrices with rank $k-1$. However, if we try to go further to $k-2 < d < k-1$, the probability density becomes negative. Even for values of $d < k-1$ at which the probability density appears to be positive, the density is not integrable near $\det M = 0$. On the other hand, since $\chi_{d,k}$ is analytic (so its Fourier transform decays exponentially) and $\chi_{d,k}(t=0) = 1$, the integral of the distribution $p_{d,k}$ over all M is well-defined and equal to unity. The resolution is that $p_{d,k}$ becomes a singular distribution which must be defined by a principal value prescription, and which is not positive definite on the singular submanifold $\det M = 0$.

As a result, the inner product on $\mathcal{H}_{\rm BU}$ can be positive definite only when all sectors with $d \notin \mathbb{N}$ have $d \geq k-1$. For a given S_∂, this requirement is most stringent for the smallest non-zero eigenvalue of Z_α, namely $d = e^{S_\partial - S_0}$. We thus find that reflection positivity can hold only when either $S_\partial - S_0$ is the logarithm of a positive integer, or $S_\partial > S_0 + \log(k-1)$.

We can use the arguments of the last section to slightly strengthen our restrictions on S_∂ by considering positivity in Hilbert spaces with boundaries, and in particular in $\mathcal{H}_{1,1}$. The discussion leading to (3.53) shows that positivity in $\mathcal{H}_{1,1}$ requires $\operatorname{rank} M \leq d$ for the matrix of inner products M in each sector $Z = d$. This is violated by the distribution (3.65) in the range $k-1 < d < k$, since M has probability density supported on matrices with full rank, $\operatorname{rank} M = k$. This gives us our final result:

$$\text{Reflection positivity} \implies e^{S_\partial - S_0} \in \mathbb{N} \text{ or } S_\partial > S_0 + \log k. \tag{3.67}$$

For any non-zero number of EOW brane species, we find that a non-zero value of S_∂ is required; the absence of a boundary action $S_\partial = 0$ does not lead to a reflection positive theory. The most natural choice is the minimal value $S_\partial = S_0$, which is the definition of the theory we used throughout the rest of this section.

The failure of models with $S_\partial = 0$ motivates us to explain the physics that might lead to an action counting the number of boundary components $|\partial M|$. This is nontrivial, because $|\partial M|$ is not a local action. For example, if we take a cylinder (with two boundaries), we can slice it in two along its length, and glue together the two edges of each piece so that we form two separate cylinders. The resulting manifold has four boundaries, so $|\partial M|$ is not preserved by this cut and paste.

However, we can achieve the same effect with a local action by introducing a new degree of freedom on each boundary. This should propagate along both asymptotic and EOW brane boundaries. Note that we regard this as part of the bulk dynamics that happens to be localised at the boundary, and not part of the dual 'CFT' dynamics. Most simply, this can be a topological quantum mechanics with Hilbert

space $\mathcal{H}_\partial$. In that case, each boundary provides a factor of $\dim \mathcal{H}_\partial$, and we can regard $-S_\partial |\partial M| = -\log \dim \mathcal{H}_\partial$ as a nonlocal effective action from integrating out this dynamics. This gives a local definition of our theory, but only if e^{S_0} is an integer. This is not entirely satisfactory: besides the somewhat artificial restriction on S_0, it seems that this degree of freedom should allow for additional boundary conditions that project onto a particular state of this boundary quantum mechanics, in which case we are again left with the theory $S_\partial = 0$.

A slightly different possibility is that some local bulk dynamics gives rise to a path integral localised at the boundary, but one which cannot be described by any quantum mechanics. This seems like a strange situation at first sight, but we note that precisely this phenomenon occurs for JT gravity. In that theory, a local bulk theory gives rise to a degree of freedom associated with asymptotic boundaries, described by the Schwarzian path integral [93, 94]. The Schwarzian alone is not a consistent quantum mechanics, since the path integral on the circle cannot be interpreted as $\operatorname{Tr} e^{-\beta H}$ for any Hamiltonian H [85, 95]. This possibility arises from a quotient by residual gauge symmetries acting nontrivially on the boundary (in that case, an $SL(2,\mathbb{R})$). Nonetheless, the gravitational theory (for example, the Lorentzian theory on a spacetime lying between two boundaries, has a good Hilbert space interpretation. While we do not have a concrete proposal to make at this time, we speculate that some analogous dynamics (or an appropriate accounting of residual gauge freedom) could naturally give rise to a theory of topology which includes a boundary effective action S_∂. In particular, we hope that our model might be obtained as a limit of a theory with more dynamics, and that this construction might offer insight into this possibility.

3.8 Spacetime 'D-branes'

We conclude the discussion of the model with some interpretative remarks for some of the results in terms of 'spacetime D-branes,' which we call SD-branes below. An SD-brane means an object on which spacetime can end, and as such is seen from spacetime as D-branes are seen from the worldsheet in string theory. In particular, they are not localised in spacetime in any way. This will be similar in spirit to the discussion of D-branes and 'eigenbranes' in [62, 82], though the framework of the Hilbert space of baby universes provides a new interpretation. We will focus on the model without EOW branes.

To study the theory in the presence of an SD-brane, we should introduce a new type of boundary of spacetime, interpreted as spacetime ending on the SD-brane. We will assign a free (possibly complex) parameter g to these boundaries, interpreted as a coupling to the SD-brane. To compute an amplitude in the presence of an SD-brane, we should allow for any number (including zero) of these additional boundaries; i.e., the

spacetime is allowed to end many times on the same SD-brane. But for the purposes of computing amplitudes, each SD-brane boundary acts much the same as a Z boundary, so we can account for them by inserting factors of gZ. To avoid overcounting different spacetimes connecting to the SD-brane, we must divide by factorials of the number of boundaries, treating the new boundaries as indistinguishable and introducing further symmetry factors where appropriate. We thus have the following recipe for computing the amplitude in the presence of an SD-brane with coupling g:

$$\begin{aligned}\left\langle f(Z)\boxed{\text{SD-brane}_g}\right\rangle &= \left\langle f(Z)\right\rangle + \left\langle f(Z)gZ\right\rangle + \left\langle f(Z)\tfrac{1}{2}(gZ)^2\right\rangle + \left\langle f(Z)\tfrac{1}{3!}(gZ)^3\right\rangle + \cdots \\ &= \left\langle f(Z)e^{gZ}\right\rangle. \end{aligned} \tag{3.68}$$

As before, the notation on the left-hand side indicates the boundary conditions for the path integral. But from the right-hand side we learn that the insertion of an SD-brane is equivalent to inserting the operator $e^{g\hat{Z}}$. In other words, the SD-brane is not a new object at all! Instead, a state $\left|\boxed{\text{SD-brane}_g}\right\rangle$ containing an SD-brane was already present in $\mathcal{H}_{\text{BU}}$ as a coherent state $\left|e^{gZ}\right\rangle$ of baby universes. We may thus identify the corresponding boundary conditions:

$$\boxed{\text{SD-brane}_g} = e^{gZ}. \tag{3.69}$$

This exponential of Z is somewhat analogous to the determinant $\det(E-H)$ introduced in [62], where it was interpreted as a brane in JT gravity. The determinant is analogous because it can be written as the exponential $\exp\left(\operatorname{Tr}\log(E-H)\right)$ of the single boundary object $\operatorname{Tr}\log(E-H)$ (single-trace in the dual matrix integral).

Now, what do the amplitudes actually look like in the presence of an SD-brane? To answer this, we compute the generating function (3.11) in an SD-brane state:

$$\begin{aligned}\left\langle\boxed{\text{SD-brane}_g}\middle|e^{uZ}\middle|\boxed{\text{SD-brane}_g}\right\rangle &= \left\langle e^{g^*Z}e^{uZ}e^{gZ}\right\rangle \\ &= \left\langle e^{(u+2\operatorname{Re}g)Z}\right\rangle \\ &= \exp\left(\lambda e^{u+2\operatorname{Re}g}\right) \\ &= \exp\left(\tilde{\lambda}e^{u}\right), \qquad \tilde{\lambda} = e^{2\operatorname{Re}g}\lambda\,. \end{aligned} \tag{3.70}$$

We here used the result $\left\langle e^{uZ}\right\rangle = \exp(\lambda e^u)$ of (3.14) in the Hartle-Hawking state, with a shifted value of u due to the presence of the SD-brane. The result (3.70) tells us is that amplitudes in the presence of an SD-brane are the same as amplitudes in the Hartle-Hawking state, but with a different value of the coupling λ. In fact, we can move between any positive real values of λ by adding an appropriate SD-brane. This is a

familiar situation from worldsheet string theory, where different values of an apparently free parameter (e.g. the coupling of the string to the Euler characteristic) turn out to describe different states of the same theory (e.g. coherent states of the dilaton).

We can also make use of these SD-branes in yet one more way by considering the effect of the imaginary part of the coupling $\theta = \mathrm{Im}\, g$. This has no effect in the amplitude (3.70), and to see its relevance we must allow for a different kind of SD-brane state in which g is not fixed but instead has a superposition of different values for θ. First, we note that the representation of the SD-brane as e^{gZ} and the integer spectrum for Z imply that θ should be understood to be periodic with period 2π. A natural basis of states superposing different values of θ is thus defined by the Fourier transformed states,

$$\left|\widetilde{\boxed{\text{SD-brane}_d}}\right\rangle := \int_{-\pi}^{\pi} \frac{d\theta}{2\pi} e^{-id\theta} \left|\boxed{\text{SD-brane}_{i\theta}}\right\rangle, \quad d \in \mathbb{N}, \tag{3.71}$$

where for simplicity we will now focus on the case $g = i\theta$, or $\mathrm{Re}\, g = 0$. In particular, the above basis diagonalizes the inner product:

$$\left\langle \widetilde{\boxed{\text{SD-brane}_{d'}}} \middle| \widetilde{\boxed{\text{SD-brane}_d}} \right\rangle = \delta_{dd'} \frac{\tilde{\lambda}^d}{d!} \qquad (d, d' \in \mathbb{N}). \tag{3.72}$$

For $d < 0$, this inner product vanishes, indicating that the resulting state is null.

To understand these states better, we may use the representation (3.69) of the SD-brane states as an exponential to write them as

$$\left|\widetilde{\boxed{\text{SD-brane}_d}}\right\rangle := \int_{-\pi}^{\pi} \frac{d\theta}{2\pi} e^{-id\theta} \left|e^{i\theta Z}\right\rangle = (-1)^d \left|\frac{\sin(\pi Z)}{\pi(Z-d)}\right\rangle. \tag{3.73}$$

But this is precisely the expression we gave in (3.25) for the α-state $\big|Z = d\big\rangle$! Furthermore, it is now clear that taking $\mathrm{Re}\, g \neq 0$ simply rescales the resulting state $\big|Z = d\big\rangle$.

This means that we can give a somewhat geometric description of a given α-sector by including a particular (Fourier transformed) $\widetilde{\text{SD}}$-brane. This $\widetilde{\text{SD}}$-brane is not a new fundamental object, but is built from a coherent state of interacting baby universes. The $\widetilde{\text{SD}}$-brane description of α-states is at first sight rather different from the alternative geometric interpretation given in section 3.3 where the $Z = d$ sector arose after constraining the path integral to spacetimes with d connected components. However, we see that the two are equivalent in the end. We expect a similar equivalence to arise in the model with EOW branes, and correspondingly in the JT gravity contexts of [62, 82].

4 Entropy bounds and the Page curve

A remarkable property of our models above was the strong role played by null states, and in particular the bound (3.41) on the rank of the inner product in any α-sector with $Z_\alpha = d$. In section 3.6 we showed this bound to follow from an abstract argument involving the cylinder state $\left|\circlearrowright\right\rangle$ in the Hilbert space $\mathcal{H}_{1,1}$ associated with a pair of disconnected boundaries. As the reader may already realize, it is straightforward to generalize this argument so as to apply to very general reflection positive gravitational path integrals. More realistic models will likely have an infinite number of states in any $\mathcal{H}_\Sigma$, so to obtain a meaningful bound on the number of states we must impose a constraint. We will achieve this here by bounding the entropy of mixed states in $\mathcal{H}_\Sigma$ with a given expected energy E.

4.1 Entropy bounds

We now state this form of the argument using the more general notation from section 2. The ideas are closely related to those in [96]. As before, we work in some definite (but arbitrary) α-sector of the given theory and also choose a spatial boundary manifold Σ; i.e., we consider a particular Hilbert space $\mathcal{H}_\Sigma^\alpha$ from section 2.4.

One property we require of our theory is that there is a notion of time evolution, here in Euclidean time. This means that the allowed boundary conditions include Euclidean 'cylindrical' boundary manifolds $C_\beta = \Sigma \times I_\beta$ for intervals I_β of arbitrary length $\beta > 0$. According to the general principles of section 2, this boundary condition describes an operator on $\mathcal{H}_\Sigma^\alpha$ that we may call $e^{-\beta H}$ and for which $e^{-\beta_1 H} e^{-\beta_1 H} = e^{-(\beta_1+\beta_2)H}$. For a given state $\left|\psi[J]\right\rangle$ defined by sources J on a boundary manifold $\mathcal{M}$ (with $\partial\mathcal{M} = \Sigma$), the action of $e^{-\beta H}$ on $\left|\psi[J]\right\rangle$ simply defines a new source J_β on a larger boundary manifold $\mathcal{M}_\beta = I_\beta \mathcal{M}$ constructed by gluing I_β to $\mathcal{M}$,

$$e^{-\beta H}\left|\psi[J]\right\rangle = \left|\psi[J_\beta]\right\rangle. \tag{4.1}$$

The final property we require of our theory is that the CPT conjugation acting on boundary conditions acts trivially on I_β. When $e^{-\beta H}$ is trace-class, this condition ensures that states $\phi_a \in \mathcal{H}_\Sigma^\alpha$ define a Hermitian matrix $(\phi_b, e^{-\beta H}\phi_a)_\alpha$ which can be diagonalized to yield discrete eigenvalues with finite degeneracy. We will take this to be the case for now and return later to the possibility that $e^{-\beta H}$ might fail to be trace-class.

The above semi-group property of $e^{-\beta H}$ then implies that the eigenvectors can be chosen to be independent of β. Together with Hermiticity, it also implies the relation

$e^{-\beta H} = \left(e^{-\beta H/2}\right)^\dagger e^{-\beta H/2}$ so that the eigenvalues must be non-negative. Henceforth, we thus take ϕ_a to denote such an orthonormal eigenbasis of $\mathcal{H}^\alpha_\Sigma$ with eigenvalues $e^{-\beta E_a}$.

The key fact is then that the boundary conditions $e^{-\beta H}$ must also define an operator on the baby universe Hilbert space $\mathcal{H}_{\rm BU}$, which we can use to define cylinder states by acting on the α-states $\Big|\alpha\Big\rangle \in \mathcal{H}_{\rm BU}$ in direct analogy with section 3.6:

$$\widehat{e^{-\beta H}}\Big|\alpha\Big\rangle = \Big|\smile_\beta;\alpha\Big\rangle \in \mathcal{H}^\alpha_{\Sigma^*\sqcup\Sigma} \quad . \tag{4.2}$$

We will be interested in forming mixed states on $\mathcal{H}^\alpha_\Sigma$, which can be thought of as elements of the Hilbert space $\mathcal{H}^\alpha_{\Sigma^*}\otimes\mathcal{H}^\alpha_\Sigma$, spanned by products $\phi^*_b\otimes\phi_a$ of our eigenstates $\phi_a \in \mathcal{H}^\alpha_\Sigma$ and their CPT conjugates. This space of density matrices is isometrically embedded via states $|\phi^*_b,\phi_a;\alpha\rangle$ into the 'two-sided Hilbert space' $\mathcal{H}^\alpha_{\Sigma^*\sqcup\Sigma}$ associated with two copies of our spatial boundary Σ. Since these latter states were built from orthonormal eigenstates of $e^{-\beta H}$ on $\mathcal{H}^\alpha_\Sigma$, the overlaps are given by

$$\Big\langle\phi^*_b,\phi_a;\alpha\Big|\smile_{\beta/2};\alpha\Big\rangle = \delta_{ab}e^{-\beta E_a/2}\,, \tag{4.3}$$

$$\Big\langle\phi^*_{b'},\phi_{a'};\alpha\Big|\phi^*_b,\phi_a;\alpha\Big\rangle = \delta_{ab'}\delta_{a'b}\,. \tag{4.4}$$

The last overlap we require is the norm of the state $\Big|\smile_{\beta/2};\alpha\Big\rangle$. This involves gluing two cylinders of length $\beta/2$ to create boundary conditions with a circle of length β: we have $\widehat{\smile_{\beta/2}}^\dagger\widehat{\smile_{\beta/2}} = \widehat{Z(\beta)}$, where the operator $\widehat{Z(\beta)}$ acting on $\mathcal{H}_{\rm BU}$ is defined by boundary conditions $\Sigma\times S^1_\beta$, with a thermal circle S^1_β of length β. The norm of our cylinder state is then given by

$$\Big\langle\smile_{\beta/2};\alpha\Big|\smile_{\beta/2};\alpha\Big\rangle = Z_\alpha(\beta), \tag{4.5}$$

where $Z_\alpha(\beta)$ is the eigenvalue of $\widehat{Z(\beta)}$ in the α state, $\widehat{Z(\beta)}\Big|\alpha\Big\rangle = Z_\alpha(\beta)\Big|\alpha\Big\rangle$.

We now introduce a state

$$\Big|\Delta\Big\rangle = \Big|\smile_{\beta/2};\alpha\Big\rangle - \sum_a e^{-\beta E_a/2}\Big|\phi^*_a,\phi_a;\alpha\Big\rangle, \tag{4.6}$$

and impose that its norm is nonnegative,

$$\langle\Delta|\Delta\rangle = Z_\alpha(\beta) - \sum_a e^{-\beta E_a} \geq 0. \tag{4.7}$$

As in section 3.6, it is important that this computation was performed in a fixed α-sector. While we arrived at (4.6) under the assumption that $e^{-\beta H}$ is trace class, a similar argument using approximate eigenvectors would in any case bound the trace of

$e^{-\beta H}$ by $Z_\alpha(\beta)$. Thus the case where $e^{-\beta H}$ fails to be trace class cannot occur and we can use (4.6) and (4.7) as written.

We can use the inequality (4.7) to make some more direct statements about the spectrum of states in $\mathcal{H}^\alpha_\Sigma$. Firstly, we can use it to bound the number of orthogonal states $N(E)$ with bounded energy $E_a \leq E$. In a thermodynamic limit, we would usually expect this to be dominated by states with energy close to the maximum, so $N(E)$ is controlled by the density of states at energy E. To bound this quantity, note that $\sum_a e^{-\beta E_a} \geq N(E) e^{-\beta E}$, by dropping all states with $E_a > E$ in the sum. From the result (4.7) we can then say that $N(E) \leq e^{\beta E} Z_\alpha(\beta)$ for any β. The sharpest bound is obtained by minimising over all β, finding

$$\log N(E) \leq S_\alpha(E), \tag{4.8}$$

where

$$S_\alpha(E) := \inf_\beta \{\beta E + \log Z_\alpha(\beta)\}. \tag{4.9}$$

This quantity is nothing but the Legendre transform of $\log Z_\alpha(\beta)$, which is the usual way of obtaining the canonical entropy from a partition function. In a semiclassical theory, and in the overwhelming majority of α-states, we expect $S_\alpha(E)$ to be approximately the Bekenstein-Hawking entropy of an appropriate black hole. This is because $Z_\alpha(\beta)$ is defined by the Gibbons-Hawking path integral with periodic Euclidean boundary conditions [61], computed semiclassically by the on-shell action of a classical Euclidean black hole. The associated entropy $S_\alpha(E)$, defined as the Legendre transform of $\log Z_\alpha(\beta)$, is then given by the Bekenstein-Hawking formula. This remains accurate in typical α states (in the measure of the Hartle-Hawking ensemble) as long as the variance of the $\widehat{Z(\beta)}$ operator is small. This is the case if connected wormhole configurations between two asymptotic $Z(\beta)$ boundaries are suppressed.

The same quantity $S_\alpha(E)$ appears in a stronger bound, constraining the von Neumann entropy $S(\rho)$ of any mixed state ρ on $\mathcal{H}^\alpha_\Sigma$. This constraint depends on the energy expectation value $E = \mathrm{Tr}(\rho H)$, where from our earlier considerations we can define H on $\mathcal{H}^\alpha_\Sigma$ by matrix elements $(\phi_b, H\phi_a)_\alpha = E_a \delta_{ab}$. Specifically, we prove that

$$S(\rho) \leq S_\alpha(E) \quad \text{for } \rho \text{ any density matrix on } \mathcal{H}^\alpha_\Sigma \text{ with } \mathrm{Tr}(\rho H) = E. \tag{4.10}$$

It suffices to show this for the density matrix that maximises $S(\rho)$ subject to the energy constraint. This is simply a Gibbs state,

$$\rho_{\mathrm{Gibbs}}(\beta) = \frac{e^{-\beta H}}{Z_{\mathrm{Gibbs}}(\beta)}, \quad Z_{\mathrm{Gibbs}}(\beta) = \mathrm{Tr}(e^{-\beta H}) = \sum_a e^{-\beta E_a}, \tag{4.11}$$

where we choose β to fix the desired energy,

$$E = -\frac{\partial}{\partial\beta}\log Z_{\text{Gibbs}}(\beta). \tag{4.12}$$

Note that Z_{Gibbs} is precisely the quantity we bounded in (4.7), with the inequality $Z_{\text{Gibbs}}(\beta) \leq Z_\alpha(\beta)$. Now, we can compute the von Neumann entropy of ρ_{Gibbs} as the Legendre transform of Z_{Gibbs}:

$$S(\rho_{\text{Gibbs}}(E)) = \inf_\beta\{\beta E + \log Z_{\text{Gibbs}}(\beta)\} \tag{4.13}$$

$$\leq S_\alpha(E) \tag{4.14}$$

The inequality follows because $S_\alpha(E)$ is defined in (4.9) by the same minimisation as used here to obtain $S(\rho_{\text{Gibbs}}(E))$, after replacing $Z_{\text{Gibbs}}(\beta)$ by the larger function $Z_\alpha(\beta)$. This demonstrates the claimed entropy bound (4.10).

4.2 Consequences and interpretations

Our results (4.8) and (4.10) show that, for theories defined by reflection positive path integrals, the density of states in any $\mathcal{H}_\Sigma^\alpha$ is bounded by $S_\alpha(E)$ from (4.9), which generically we expect to be given by the Bekenstein-Hawking entropy of an appropriate black hole.

We interpret this result as a semiclassical Page curve. The class of mixed states ρ on $\mathcal{H}_\Sigma^\alpha$ that we can prepare by asymptotic sources includes old black holes. For example, we can create pure state black holes by collapse, couple to an auxiliary 'bath' system into which the Hawking radiation escapes, and trace out the bath. In the usual semiclassical description, it seems that this process can produce states of a given energy with arbitrarily large entropy. This entropy comes from the large interior which grows with time (in particular linearly with time along a 'nice slice' [97]), which can be populated with a growing number number of naively distinct possible low energy states. Our result shows that in an alpha sector of a reflection positive path integral, nonperturbative effects giving exponentially small overlaps between these states must conspire to produce surprising linear relations between them. Such relations must occur after the Page time so that the entropy of the black hole is bounded by the Bekenstein-Hawking entropy, to satisfy (4.10). If this inequality is (approximately) saturated, the entropy of the black hole (i.e. the density matrix on $\mathcal{H}_\Sigma^\alpha$) and of the radiation will follow the Page curve.

We expect that in contexts where the naive number of states in $\mathcal{H}_\Sigma$ can be made arbitrarily large, one will find that the bound $S(\rho) \leq S_\alpha(E)$ of (4.10) can be saturated, as in our model with large k. In particular, we expect this to hold for the old black

holes in the discussion above. This requires saturation of the inequality in (4.7) for all β, and so $|\Delta\rangle$ becomes a null state. Note that $|\Delta\rangle = 0$ is equivalent to the statement that $Z_\alpha(\beta)$ is equal to the actual thermal partition function $\mathrm{Tr}\, e^{-\beta H}$ on $\mathcal{H}_{\Sigma,\alpha}$. The result that the function $Z_\alpha(\beta)$ can be written as a thermal trace is a strong constraint on the eigenvalues of $\widehat{Z(\beta)}$, which should be viewed as generalizing the result $Z_\alpha \in \mathbb{N}$ from our models in section 3.

In the case of saturation, the statement that $|\Delta\rangle$ is null leads to a gauge equivalence

$$\Big|\smile_{\beta/2}; \alpha\Big\rangle = \sum_a e^{-\beta E_a/2} \big|\phi_a^*, \phi_a; \alpha\big\rangle. \tag{4.15}$$

Following [11], the cylinder state is naturally associated with a two-sided black hole with an Einstein-Rosen bridge joining the two boundaries. We see the familiar equivalence between this and a superposition of product states emerging as an example of our gauge equivalence.

To connect further with our desire to understand black hole evaporation, we recall from section 2 that for any state ρ prepared with asymptotic sources, the Rényi (and von Neumann) entropies $S_n(\rho)$ of ρ again define operators on $\mathcal{H}_{\rm BU}$ and take definite values in α-sectors. These entropies are then subject to versions of the above bound in each α-sector, and as a result so are their expectation values $\Big\langle S_n(\rho) \Big\rangle$ in the Hartle-Hawking state. In the context of black holes, any such entropies will then reproduce an appropriate Page curve defined by the Bekenstein-Hawking entropy. In particular, the final result will then be much as in the recent discussions of replica wormholes [48, 49] which in our language are indeed the most natural saddle points contributing to the average entropy $\Big\langle S_n(\rho) \Big\rangle$.[14] The argument above shows that similar results will then hold when one computes the full result of any reflection positive gravitational path integral. Further, it tells us that these bounds hold not just on average, but in every α-state. This puts additional constraints on higher moments of the entropy.

It is, however, important to note the precise sense in which the entropies $\Big\langle S_n(\rho) \Big\rangle$ have just been defined. From our perspective, the basic quantities are the eigenvalues $S_{n,\alpha}(\rho)$ of $\widehat{S_n(\rho)}$ in the various α-states. These are entropies defined separately on each $\mathcal{H}_{\Sigma,\alpha}$. Working in the Hartle-Hawking state then computes the average $\Big\langle S_n(\rho) \Big\rangle$ of such entropies over the α-states in the Hartle-Hawking ensemble. In particular, while this $\Big\langle S_n(\rho) \Big\rangle$ is computed by replica wormholes (to a first approximation), it manifestly does *not* include entanglement with the baby universe sector.

[14] More properly, the replica wormholes are saddle points for $\Big\langle \mathrm{Tr}(\rho^n) \Big\rangle$, but the distinction is unimportant as long as the variance of these quantities is small.

This is a physically useful notion of entropy as the α-sectors are superselected from the standpoint of asymptotic observers, and entanglement with superselection sectors is in principle unobservable. Nevertheless, if one wishes to consider the entropy of some density matrix on the full space $\mathcal{H}_\Sigma$ (and not just on a single α-sector) defined by some fixed set of sources, entanglement with baby universes will generally lead to much larger entropies that exceed the Bekenstein-Hawking entropy and thus do not reproduce the expected Page curve. In this more mathematical sense, Hawking was correct [98] that information is lost in black hole evaporation. This is all in direct parallel with the conclusions of [55–57, 99] from long ago. We will also discuss such connections in more detail in a forthcoming companion paper.

5 On third-quantized perturbation theory

5.1 Formulating a wormhole perturbation theory

We have been interested above in contexts where spacetime wormholes provide the dominant effects. But in most circumstances spacetime wormholes are not the minimum action configurations. In such cases, it is natural to expect other configurations to dominate, and for the contributions of spacetime wormholes to be nonperturbatively suppressed by a factor of the form e^{-S}, where S, of order G_N^{-1}, is the action of a wormhole. This holds for computing simple amplitudes in our models of section 3, for which higher topologies are suppressed by factors of the large parameter λ. In such cases it is natural to use an approximation where different universes evolve independently at leading order, and where spacetime wormholes are included as perturbative interactions between universes. The resulting perturbation theory is the 'third quantised' formalism of [57]. This approximation was also emphasized in other contemporaneous literature on wormholes [55, 56, 99, 100].

We now describe an analogous approximation in our framework. This will serve both to complete the connection to the above literature and to provide a better understanding of the interesting circumstances described above in which this approximation fails. Nevertheless, this section represents a distraction from the main line of inquiry presented here, and some readers may wish to skip directly to section 6.

The early works [55–57] focused on studying microscopic wormholes, with the intent of describing physics on distances scales much larger than the wormhole's characteristic size (say, Planck scale). The relevant scale is the 'width' of the wormhole mouth, thought of as some length scale associated with the cross-sectional area. In contrast, the separation between the spacetime regions associated with the wormhole mouths can be much larger. In that context, it is most natural to describe the physics using

the operators of the low energy effective field theory, studying the effect of integrating out the microscopic wormholes. In contrast, we have wormhole mouths which, as with replica wormholes, are determined by a classical or quantum extremal surface. As a result, our wormholes will typically have a size similar to some black hole horizon, which may be both macroscopic and large. For us it thus will be more natural to discuss CFT boundary operators $\widehat{Z[J]}$ in place of the low energy bulk fields. This captures much of the same physics, and is analogous to using an S-matrix description in place of an effective Lagrangian.[15] The effects on the bulk effective field theory that arise from integrating out macroscopic wormholes will be explored in section 6.

Suppose then that, for some theory and amplitude of interest, the contribution from topologies connecting many boundaries is suppressed relative to disconnected topologies. This holds for familiar simple amplitudes in theories of interest, including the model discussed in section 3, as well as for JT gravity — though it does not hold for all amplitudes, as we will discuss below. In a case where it does, at zeroth order of approximation we may neglect the connected contributions, obtaining an amplitude that approximately factorizes:

$$\mathfrak{Z}^{-1}\Big\langle Z[J_1]\cdots Z[J_n]\Big\rangle \approx \mathfrak{Z}^{-n}\Big\langle Z[J_1]\Big\rangle\cdots\Big\langle Z[J_n]\Big\rangle \tag{5.1}$$

Identifying an asymptotically AdS boundary $Z[J]$ with an operator $\widehat{Z[J]}$ acting on the baby universe Hilbert space $\mathcal{H}_{\rm BU}$ as in (2.11), at this leading order of approximation we can simply replace $\widehat{Z[J]}$ with a multiple of the identity operator $\mathfrak{Z}^{-1}\langle Z[J]\rangle$. In particular, at this level of approximation, acting with any $\widehat{Z[J]}$ on $\big|\mathrm{HH}\big\rangle$ yields another state proportional to $\big|\mathrm{HH}\big\rangle$, so the baby universe Hilbert space defined in section 2.2 collapses to a single dimension.

To incorporate nontrivial wormhole physics, we must go to next order in the approximation, allowing contributions to the path integral from spacetimes that connect either one or two asymptotic boundaries, but not more. The contributions from spacetimes with one asymptotic boundary are then analogous to quantum field theory tadpoles, while the two boundary contributions are analogous to quantum field theory propagators. In particular, the Hilbert space $\mathcal{H}_{\rm BU}$ becomes nontrivial, and takes the form of a Fock space. To see this, we define 'single universe states' by subtracting the 'tadpole contributions' from one boundary states; i.e., one need only introduce the modified (tilded) states

$$\widetilde{|Z[J]\rangle} = |Z[J]\rangle - \mathfrak{Z}^{-1}\langle Z[J]\rangle|\mathrm{HH}\rangle, \tag{5.2}$$

[15] In the language of [101], the effects of higher topology we study are more closely analogous to 'wormhole interactions', as opposed to the 'instanton interactions' arising from nearby wormhole mouths of primary interest in that work.

and similarly for states involving larger numbers of universes. Loosely speaking, the spacetime created by the operator $\widehat{Z[J]}$ is most likely to immediately cap off, failing to create a closed universe. It is natural to subtract this possibility, in which case we are most likely to create a single closed universe which can propagate to another asymptotic boundary, justifying the name of 'single universe state'. Going to higher orders in the approximation would require additional subtractions for this description to remain valid.

The resulting Fock space structure can be used to define baby universe creation and annihilation operators $a_J^\dagger, a_{J^*}$, where in particular we have

$$a_J|\mathrm{HH}\rangle = 0; \tag{5.3}$$

$$a_J^\dagger|\mathrm{HH}\rangle = |Z[J]\rangle - \mathfrak{Z}^{-1}\langle Z[J]\rangle|\mathrm{HH}\rangle, \tag{5.4}$$

and the algebra $[a_{J_1}, a_{J_2}] = 0$,

$$\Big[a_{J_1}, a_{J_2}^\dagger\Big] = \Big\langle Z[J_1^*]Z[J_2]\Big\rangle - \mathfrak{Z}^{-1}\Big\langle Z[J_1^*]\Big\rangle\Big\langle Z[J_2]\Big\rangle. \tag{5.5}$$

One can then write corrections to the boundary operators $\widehat{Z[J]}$ in terms of baby universe creation and annihilation operators:

$$\widehat{Z[J]} \sim \mathfrak{Z}^{-1}\big\langle Z[J]\big\rangle \ + \ a_J^\dagger + a_{J^*} + \cdots\,, \tag{5.6}$$

where $\cdots$ indicates higher order terms.

One is then tempted to think of the states $\widetilde{|Z[J]\rangle}$ as (approximations to) states of a single closed baby universe, with a wavefunction for the metric and other fields determined by the source J (and by varying J we would expect to obtain an overcomplete set of coherent states). We can diagonalise the inner product on the single-universe Hilbert space, taking linear combinations of $\widehat{Z[J]}$ for different J to give operators $\widehat{Z}_i$ which are chosen to be Hermitian and give amplitudes satisfying

$$\mathfrak{Z}^{-1}\langle Z_i Z_j\rangle - \mathfrak{Z}^{-2}\langle Z_i\rangle\langle Z_j\rangle = \delta_{ij}\,. \tag{5.7}$$

We can then write $\widehat{Z}_i = \langle Z_i\rangle + a_i^\dagger + a_i + \cdots$, with a more conventional oscillator algebra $[a_i, a_j^\dagger] = \delta_{ij}$ labelled by an orthonormal basis of single-universe states. Repeated applications of $a_i^\dagger$ are then said to create more universes, which can interact through topologies connecting three or more boundaries and into which we could incorporate as higher order terms in (5.6). As long as these higher topologies are suppressed, we can thus construct a useful perturbation theory, where the inner product in (5.5) gives the 'free propagator' for single universe states, with higher topologies contributing vertices.

In particular, as noted above, based on the validity of the free approximation $\mathcal{H}_{\rm BU}$ appears to be well described by a Bosonic Fock space built on the single-universe Hilbert space. The Hartle-Hawking state provides the oscillator ground state, and multi-universe states are built by acting with $a_i^\dagger$ operators. Alternatively, in the free approximation we can think of $\mathcal{H}_{\rm BU}$ in terms of the wavefunction $\Psi(Z_i)$, a function of the real variables Z_i. The operator $\widehat{Z}_i$ then acts as a position operator (or a free field operator in QFT, where the label i could be momentum, for example), multiplying by Z_i. As the oscillator vacuum, the Hartle-Hawking state has a Gaussian wavefunction for each Z_i, shifted to be centred on $\langle Z_i\rangle$.

It is now tempting to use this free Fock space description to describe the spectrum of $\widehat{Z[J]}$, and hence the dual ensemble and the α-states. We are led to expect that the spectrum of $\{Z_i\}$ has continuous support on the whole of $\mathbb{R}$, independently for every i. In the resulting ensemble the Z_i, and hence the $Z[J]$, are normally distributed at the first nontrivial order described above, with covariance matrix given by the single-universe inner product[16] in (5.5). At each higher order, corrections from interactions would then appear to contribute only small non-Gaussian corrections to the measure, the conclusion reached in [101], for example. However, in this respect, we have been misled by the free 'approximation' 5.6. It turns out to be invalid because, while perturbation theory is accurate in many circumstances, it is not applicable in α-states, as we will argue in a moment. The true, nonperturbative spectrum is smaller because the Fock space description of the Hilbert space is invalid once we take into account the null states (2.9) by which we must quotient by to obtain $\mathcal{H}_{\rm BU}$. Due to the null states, the 'universe number' which grades the Fock space is not a diffeomorphism invariant observable.

Before we describe the breakdown of third-quantised perturbation theory, we clarify that it is not necessarily signalled by the dominance of spacetime wormhole effects. It may happen that the most important contribution to an amplitude comes from a nontrivial topology, but higher topologies remain negligible. This occurs prominently in two recent examples. The first is the spectral form factor $\langle Z(\beta+it)Z(\beta-it)\rangle$ of JT gravity [62, 86, 102], for which the contribution from the disconnected topology decays in time, while the connected topology gives a contributions that is exponentially suppressed but growing. Eventually, the connected topology dominates, giving the 'ramp'. A second example is the nth Rényi entropy of an evaporating black hole after the Page time, which can be described as a sum of n-boundary amplitudes; the dominant configuration is a 'replica wormhole', a spacetime which connects the n boundaries

[16] This is equivalent to the statement that the vacuum state of a free field theory is Gaussian with corresponding covariance matrix.

[48, 49]. However, higher topologies continue to be suppressed in such cases, and a similar perturbation theory remains valid; it simply happens to be dominated by n-universe vertices, so requires their inclusion.[17]

Instead, we are interested in cases when the third quantised perturbation theory fails entirely, and many topologies must be considered at once. For example, this occurs when we compute amplitudes with a parametrically large number of boundary components, giving very large moments of $\widehat{Z[J]}$. Equivalently, we can describe these amplitudes as the overlaps of states with very large universe occupation number[18]. While any particular process of splitting and joining universes is suppressed, the total amplitude of such interactions is enhanced by combinatorial factors counting the number of processes with many possible universes (or joining many possible boundaries). This allows higher topologies to become important.

Crucially, this breakdown of perturbation theory applies to α-states and so is vitally important for understanding the spectrum of $\widehat{Z[J]}$. The approximation of weakly interacting baby universes is thus not a reliable guide to the details of the spectrum. In the free theory, the α-states are like position eigenstates in the harmonic oscillator. They thus have infinite expectation value for the number operator. As we reduce the uncertainty in the α parameters and create a baby universe wavefunction with a more narrow spread, the mean universe occupation number increases, and eventually becomes exponentially large. At that point, the above approximation is not self-consistent for studying such states.

In retrospect, it should not be surprising that perturbation theory is of limited use for determining the spectrum of observables. As a simple example of similar behavior, if we perturb around the minimum of a potential in quantum mechanics, we cannot at any finite order tell whether the configuration space is compact, and hence if the momentum should be quantised.[19]

The truncation of the spectrum of $\widehat{Z[J]}$ is invisible at any finite order in the third-quantised perturbation theory. Thus in that description it could be seen only via some nonperturbative effect, or in an exact solution if one turns out to be available. Our models of 3 provide a simple example of the latter. Recall that, in terms of the usual bulk perturbation theory in G_N, the spacetime wormholes describing third-quantised

[17] This perturbation theory is also useful for discussing the average entanglement spectrum close to the Page time [49], though it requires summation of a class of 'tree-level' diagrams involving vertices of all valences.

[18] This notion is well-defined only in the third quantised perturbation theory, but can nonetheless be used to diagnose whether that perturbation theory is self-consistent.

[19] We mentioned above the natural third quantization interpretation of $\widehat{Z[J]}$ as a position-like operator, but we could equally well have interpreted it as an analogue of free particle momentum

interactions are already nonperturbative, so the relevant expansion parameter is of the form e^{-S} for an action S of order G_N^{-1}. From this point of view, the compression of the Hilbert space is then a *doubly* nonperturbative effect, contributing to simple amplitudes as $e^{-c\,e^{-S}}$ for some (possibly imaginary) constant c.

5.2 Perturbation theory in the topological model

To give some insight into the validity of third quantised perturbation theory, we discuss its applicability in the context of the model of section 3. We will restrict our considerations to the model without EOW branes.

The small parameter that suppresses topology is e^{-S_0}, with S_0 multiplying the Euler characteristic. It is natural to organise the third quantised perturbation theory as an expansion in that parameter, with higher genus topologies appearing as loops. However, the details of such an expansion (particularly accounting for diffeomorphisms of connected surfaces) are not necessary for the point we wish to illustrate. To simplify the discussion, we thus instead assume that the full connected correlators (and thus any sums over connected surfaces with given boundaries) have already been computed exactly. These are all given by the same number λ, so our perturbation theory will be an expansion in inverse powers of λ. As noted in section 3, this expansion is organised by counting the number of connected components of spacetime.

Let us begin by noting a precise sense in which the free Gaussian approximation is appropriate at large λ. This follows from first observing that a sum of N independent Poisson distributions with parameter λ/N is again a Poisson distribution, with parameter λ. Taking λ and N large with fixed ratio then implies that we can apply the central limit theorem to the Poisson distribution as $\lambda \to \infty$. Specifically, we may define

$$X = \frac{Z - \lambda}{\sqrt{2\lambda}}, \tag{5.8}$$

which has mean zero and variance unity. This X is just new encoding of the boundary condition Z, with the shift by λ acting to subtract the 'tadpole' and set $\langle X \rangle = 0$, and with an additional rescaling to fix the variance $\mathfrak{Z}^{-1}\langle X^2 \rangle = \frac{1}{2}$. The central limit theorem then implies that as $\lambda \to \infty$ the distribution of X converges to a normal (and thus Gaussian) distribution. In particular, at large λ any amplitudes $\langle f(X) \rangle$ for bounded continuous functions f (fixed independently of λ) approach those computed by integrating against a Gaussian. These are the vacuum amplitudes of a harmonic oscillator, with wavefunction $\propto e^{-\frac{x^2}{2}}$, so this defines the 'free' Gaussian approximation mentioned above.

We will return to the discussion of this wavefunction later. Before doing so, we the large λ expansion to study the moments $\mathfrak{Z}^{-1}\langle Z^n \rangle = B_n(\lambda)$ and note both when and

how that expansion fails as we also take n to be large. For fixed n, the leading order contribution at large λ comes from completely disconnected spacetimes, giving $B_n(\lambda) \sim \lambda^n$. At the next order, we have spacetimes with $n-1$ disconnected components, which requires one 'cylinder', a component joining two boundaries.[20] There are $\binom{n}{2} = \frac{n(n-1)}{2}$ choices of which boundaries to join, so we have

$$B_n(\lambda) = \lambda^n + \frac{n(n-1)}{2}\lambda^{n-1} + \cdots \quad \lambda \to \infty, \text{ fixed } n. \tag{5.9}$$

At the next order, we have spacetimes with $n-2$ components, which means either two cylinders, or a 'pair of pants' connecting a trio of boundaries to the same component of spacetime. We can continue in this way to any desired order λ^{n-k} in the expansion by accounting for possible topologies with $n-k$ connected components.

Now, let us consider what happens when n also becomes large. The first sign of trouble occurs when n if of order $\sqrt{\lambda}$, when the second term in the above expansion is no longer smaller than the first. There are roughly $n^2/2$ ways to choose pairs of boundaries to join by a cylinder (neglecting the correction from choosing the same boundary twice), which is sufficiently large to overcome the suppression by λ. But this does not apply only for a single cylinder; terms with any number of cylinder components again contribute at the same (leading) order. In some sense our free approximation has failed.

However, it turns out that the large λ expansion remains useful because we can explicitly account for the sum over configurations with k cylinder components. For $2k \ll n$, there are approximately $\frac{1}{k!}\left(\frac{n^2}{2}\right)^k$ ways to select k pairs of boundaries to join with a cylinder, where we have neglected the correction from 'interactions', where the same boundary is chosen more than once. Summing over this 'free gas of cylinders' gives us a multiplicative correction to the nth moment of Z,

$$B_n(\lambda) \sim \lambda^n e^{\frac{n^2}{2\lambda}} \quad \lambda, n \to \infty, \text{ fixed } \frac{n^2}{\lambda}. \tag{5.10}$$

In this regime, we can now systematically correct (5.10) in powers of λ^{-1} as before. Such corrections can account for including higher topologies with more boundaries as well as compensating for the overcounting of cylinder configurations.

From (5.10), we see that $\langle Z^n \rangle$ is dominated by contributions with roughly $\frac{n^2}{\lambda}$ cylinder components. This can be much greater than one and the analysis will remain applicable, though it should certainly remain much less than n, so we must have $n \ll \lambda$.

[20] For simplicity of language, we will call this a cylinder even though it packages a sum over surfaces of all genus with two boundaries. A more precise language might refer to it as a renormalized cylinder.

If this is the case, the correction from the cylinders is small in the sense that it is subleading to the λ^n term when expressed as an expansion of $\log B_n(\lambda)$.

Taking n larger still, (5.10) remains accurate until n is of order $\lambda^{2/3}$. At that point we find significant corrections from including any number of connected components having three boundaries each ('pairs of pants'), and also from certain aspects of the overcounting of configurations of multiple cylinders. In the latter context, the relevant configurations are those in which two cylinders end on the same boundary. We previously included these configurations for simplicity (and to obtain a definite power of λ), but since they are not allowed we must now compensate by subtracting off their contributions. Together, these two effects multiply (5.10) an extra factor of $e^{-\frac{n^3}{3\lambda^2}}$. This pattern continues, with similar $e^{\#\frac{n^k}{\lambda^{k-1}}}$ corrections appearing whenever n becomes of order $\lambda^{1-\frac{1}{k}}$ for $k = 2, 3, 4, \ldots$. As discussed in appendix A.2, this structure is also apparent from a direct asymptotic expansion of $B_n(\lambda)$.

In summary, in the regime $\lambda \ll n$ the large λ expansion remains a tractable way to compute the moments $\langle Z^n \rangle$ and is organized by types of contributing geometries. However, once n is of order λ, this perturbation theory breaks down catastrophically, since there is no longer any suppression of connected topologies with many boundaries. This is the regime in which the novel effects of null states and gauge invariance become relevant, truncating the spectrum of Z and making its discreteness apparent.

To explain this last statement in more detail, we first describe the state $|Z^n\rangle$ in the free approximation. We begin by translating to the harmonic oscillator position variable variable X introduced in (5.8), writing $Z^n = \lambda^n \left(1 + \sqrt{\frac{2}{\lambda}} X\right)^n$. Expanding $\log Z^n$ at large λ (but any fixed n), this gives $\log Z^n = n \log \lambda + \sqrt{\frac{2}{\lambda}} nX + O(n\lambda^{-1})$. We may thus approximate $Z^n \sim \lambda^n \exp\left(\sqrt{\frac{2}{\lambda}} nX\right)$. For sufficiently small n that the free approximation is applicable, we therefore have an approximate equivalence between the following states:

$$\left|Z^n\right\rangle \simeq \mathfrak{Z}^{1/2} \lambda^n e^{\sqrt{\frac{2}{\lambda}} n\hat{X}} \left|0\right\rangle \simeq (e\lambda)^n \left|e^{\frac{n}{\lambda} Z}\right\rangle \tag{5.11}$$

Here the final equality uses (5.8), and the middle state lives in the harmonic oscillator Hilbert space of the free approximation. In particular, $|0\rangle$ is the (normalized) oscillator vacuum with wavefunction $\psi(X) \propto e^{-\frac{X^2}{2}}$. After applying the exponential operator, the resulting wavefunction is a shifted Gaussian, which is a coherent state of the harmonic oscillator with average occupation number (here, 'universe number') $\frac{n^2}{\lambda}$. From the above analysis, it follows that the free approximation is valid for universe numbers $N \ll \lambda$.

Now, a wavefunction of width ΔX in the X variable has an occupation number that scales as $N \simeq (\Delta X)^{-2}$ as the width goes to zero, where the leading contribution comes from writing occupation number in terms of the Harmonic oscillator Hamiltonian and focusing on the kinetic term. In terms of the width ΔZ in Z, this is $N \simeq \lambda(\Delta Z)^{-2}$. But resolving the natural integer discreteness in the spectum of Z requires $\Delta Z \sim 1$, and hence N of order λ. As a result, and as one might expect, the discreteness of the Z spectrum is thus associated with the complete breakdown of third quantised perturbation theory.

We can also see directly that this regime is connected with the appearance of null states, and thus the appearance of new gauge equivalences. Perhaps the simplest equivalence is that between the Hartle-Hawking state and the exponential $\left|e^{2\pi i Z}\right\rangle$. Note that any state $\left|e^{\alpha Z}\right\rangle$ is described in the free approximation by a coherent state with average occupation number $N \sim |\alpha|^2\lambda$. But for α of order one (for example, for $\alpha = 2\pi i$) this is of order λ and the free approximation fails.

All these phenomena occur when the state of baby universes has unsuppressed interactions with a given boundary. Roughly speaking, if we have a state of $\mathcal{H}_{\rm BU}$ containing N closed universes and introduce a new boundary, the new boundary will connect to any given universe with amplitude λ^{-1}. Hence it will connect to *some* universe with amplitude N/λ. This effect becomes of leading order at N of order λ, when the free description breaks down. We emphasise that this heuristic is appropriate for $N \ll \lambda$ when the free approximation can be used, but that N itself becomes ill-defined once it becomes of order λ. At that point, null states appear and, furthermore, the null states are not preserved by any notion of universe number operator $\hat{N}$.

6 Discussion

As with many works motivated by the black hole information problem, various readers may wish to focus on either the technical aspects of the above results or, alternatively, on their further significance for quantum gravity. For this reason, we separate our discussion below into more technical remarks in section 6.1 and a broader consideration of implications in section 6.2

6.1 Summary and future directions

We have seen that combining features of AdS asymptotics with the basic perspective of Coleman [55] and of Giddings and Strominger [56, 57] from the late 1980's leads to a sharp structure in which states in a 'baby universe Hilbert space' $\mathcal{H}_{\rm BU}$ control an ensemble of results for quantities $Z[J]$ computed at asymptotically AdS boundaries.

This version of the argument uses only manifest properties of the path integral and makes no further assumptions about locality.

Nevertheless, the final result is much the same as in [55, 56]. In particular, the full bulk theory naturally includes both $\mathcal{H}_{\rm BU}$ and what one may call asymptotically AdS states, and there is a sense in which the two sectors interact. However, the theory has superselection sectors for the algebra of operators on the asymptotically AdS states, so that an observer with no access to $\mathcal{H}_{\rm BU}$ naturally experiences an ensemble. The superselection sectors are associated with a complete orthonormal basis $\{|\alpha\rangle\}$ of $\mathcal{H}_{\rm BU}$ in which the $Z[J]$ take definite values and exhibit factorization. Thus for a given state $|\Psi\rangle \in \mathcal{H}_{\rm BU}$, the probability of outcome $Z_\alpha[J]$ is $p_\alpha = |\langle\Psi|\alpha\rangle|^2$. Furthermore, all properties of the full spectrum of superselection sectors can at least in principle be computed from correlators in the Hartle-Hawking no-boundary state $|{\rm HH}\rangle \in \mathcal{H}_{\rm BU}$.

We then explored this construction in detail in simple topological models inspired by Jackiw-Teitelboim gravity with and without end-of-the-world branes (EOW branes, see e.g. [49, 103]), and perhaps also with an extra boundary degree of freedom. Without EOW branes, there is a single asymptotically AdS boundary condition Z, for which the associated operator $\widehat{Z}$ is naturally interpreted as the dimension of the CFT Hilbert space. This operator is also present in the model with EOW branes. Interestingly, the models predict this operator to have a quantized spectrum with eigenvalues $Z_\alpha \in e^{S_\partial - S_0}\mathbb{N}$, where S_∂ is a parameter associated with the extra boundary degree of freedom. The potential eigenstates associated with other potential eigenvalues turn out to be null states. Perhaps even more intriguingly, unless S_∂ is taken to be larger than $S_0 + \log k$, the models with EOW branes are reflection positive only when all Z_α are nonnegative integers, and thus only when $e^{S_\partial - S_0} \in \mathbb{N}$. The particular ensemble defined by the Hartle-Hawking no-boundary state gives a Poisson distribution for the Z_α.

Models with EOW branes have additional boundary conditions (ψ_j, ψ_i) for $i, j = 1, \ldots k$. The (ψ_j, ψ_i) are naturally interpreted as the matrix of inner products between EOW brane states in a dual boundary quantum mechanics. For given (integer) Z_α, the eigenvalues of $\widehat{(\psi_j, \psi_i)}$ take the form $\sum_a \bar{\psi}_j^a \psi_i^a$ for some rectangular matrix ψ_i^a of size $k \times Z_{\alpha_k}$. As a result, the rank of any $(\psi_j, \psi_i)_\alpha$ cannot exceed either k or Z_α. The ensemble defined by the Hartle-Hawking no-boundary state arises from choosing independent complex Gaussian random entries for each of the ψ_i^a.

For $k \gg Z_\alpha$, this structure $(\psi_j, \psi_i)_\alpha = \sum_a \bar{\psi}_j^a \psi_i^a$ requires a sizeable compression of the naive the CFT Hilbert space (which would have had dimension k). In particular, any list of more than Z_α states in the CFT Hilbert space turns out to be linearly dependent due to the presence of null states. We also argued that a similar constraint on the number of linearly dependent states must arise in any theory where the gravitational

path integral defines a positive semi-definite physical inner product. Our general argument is closely related to ideas in [96], and various related suggestions can be found in e.g. [104–108]. But the result is deeply related to recent successes [42, 43, 48, 49] in reproducing various forms of the Page curve associated with the black hole information problem. With hindsight one can say that it was implicit in all of these works, and in fact moderately explicit in [49]. But here we see that it is an exact statement at finite Z in every possible baby universe state.

Indeed, in order to explain the Rényi computations of [49] for typical members of the Hartle-Hawking ensemble some version of this compression must occur whenever the number of a priori independent states inside a quantum extremal surface exceeds the generalized entropy defined by the region outside. And due to a maximin argument [42, 43], one expects this to occur whenever the number of a priori independent quantum states that can exist inside a given bulk domain of dependence with fixed exterior geometry exceeds the area of the codimension-2 surface where the past and future boundaries of this domain of dependence intersect; see also [109] for more on quantum maximin surfaces.

In the context of black hole evaporation, for general baby universe states $\big|\Psi\big\rangle$ this picture gives a sense in which interactions with baby universes formally lead to loss of information during the evaporation of black holes. But as described previously in [55–57, 99], since the α-states define superselection sectors for asymptotic observers, any given asymptotic observer can find no operational signs of this information loss. In particular, while the observer may not be able to predict the exact outcome of an experiment involving black holes, they may simply consider the experiment to be a partial measurement of the previously unknown value of (in this interpretation unique) value of α describing the universe in which they live. To the extent that α has been measured, no further information is then lost.

At the technical level there remain many interesting generalizations to explore in the future. For example, even in the models discussed here, it would be useful to understand if one can formulate the Hilbert spaces $\mathcal{H}_{\rm BU}$ using slices at 'finite time', or in other words without reference to asymptotic boundaries. Moving beyond the current model, one would like to add topological matter, and also to explore a similarly topological version of the de Sitter models of [110] and [49]. Work along these lines is in progress and we hope to report soon. In the longer term, it is also clearly of interest to study more realistic models.

6.2 Transcending the ensemble: implications and interpretations for each α-sector

We now turn to more speculative comments concerning the implications of our results above.

A key lesson from this work appears to be that, at least in sufficiently simple models, gravitational path integrals by themselves succeed in describing a great deal of microscopic information. In particular, in our models the bulk path integral leads to a definite construction of the possible boundary theories — defined by simultaneous eigenvalues $Z_\alpha[J]$ — and also of the ensemble defined by the Hartle-Hawking state. However, this was possible only due to the exact solubility of the model, and in particular the convergence of the sum over topologies. In more realistic models, we will surely not be so fortunate.

Even in the simple case of JT gravity and its cousins [49, 62, 87], the gravitational path integral fails to converge. Though the model is sufficiently simple that the path integral for any given topology is exactly computable, the sum over topologies is an asymptotic series with zero radius of convergence in the expansion parameter e^{-S_0}. While there is an extremely natural completion of the model defined by a dual double-scaled matrix integral, it remains unclear whether the gravitational path integral uniquely selects this completion, or how it is realised in the bulk. This completion is associated with nonperturbative effects in the sum over topologies, which are *doubly* nonperturbative in G_N. The same doubly nonperturbative scale was associated with truncation of the baby universe Hilbert space in our model, suggesting a tantalising connection to explore in more generality.

If we apply the ideas of this paper to more conventional 'top-down' examples of AdS/CFT duality, such as type IIB supergravity (or string theory) with $\mathrm{AdS}_5 \times S^5$ boundary conditions, there are several possible outcomes. The first possibility, suggested by our simple model and JT gravity, is that a nonperturbatively complete bulk theory defines a large Hilbert space $\mathcal{H}_{\mathrm{BU}}$ of baby universes. The eigenstates $\big|\alpha\big\rangle$ would then be associated with a menagerie of dual CFTs, and the Hartle-Hawking state again defines an ensemble of them. However, this is in tension with the established statement of the duality, which uniquely selects $\mathcal{N} = 4$ Yang-Mills theory as a CFT dual.[21] A nontrivial ensemble would require surprising new families of maximally supersymmetric CFTs; in particular, since $\mathcal{N} = 4$ Yang-Mills is the unique such theory at weak

[21] Recall that a given α-state determines partition functions for all possible boundary conditions on the bulk fields. These boundary conditions include specifications the flux on S^5 and the asymptotic dilaton, associated with the rank N of the dual $U(N)$ gauge group and the 't Hooft coupling λ respectively. An α-state would specify a family of theories labelled by these parameters.

coupling, these new CFTs must be strongly coupled throughout their moduli space.

Perhaps the more likely scenario is that $\mathcal{N} = 4$ Yang-Mills is the unique dual and there is no ensemble. The baby universe Hilbert space interpretation is that $\mathcal{H}_{\rm BU}$ is one-dimensional, so the Hartle-Hawking state is the unique state of closed universes. The nonperturbative diffeomorphism invariance that produced null states is then required to act in the most emphatic possible fashion, rendering every possible state gauge equivalent. This unique state must then also be an α-state, and must exhibit factorization despite the existence of spacetime wormholes. Nevertheless, in analogy with typical α-states in our model, it remains possible that simple spacetime wormhole configurations still give excellent approximations to certain amplitudes. Of course, in analogy with highly atypical α-states in our model, it is also possible that that simple spacetime wormhole configurations always receive large corrections.

An intermediate position is that the bulk theory leads to an ensemble interpretation in an asymptotic (say, large N) expansion, but there is a unique theory at any finite N. This is consistent with the observation [111] that essentially any effective field theory in AdS solves the bootstrap order by order in large N perturbation theory. We can thus emulate a consistent CFT in a large N expansion, which nevertheless need not exist at any given finite N.

In any case, the suggestion is that the gravitational path integral should contain the full physics in each consistent α-sector. And since the baby universe state in such sectors does not change, there is no room in a given sector for information loss. As a result, the gravitational path integral should teach us how each consistent α-sector transfers information to the outgoing Hawking radiation.

With this in mind, we recall that a key feature of the discussion in [55–57] was the idea that one could integrate out the spacetime wormholes and describe their effects in terms of a modified effective action in which the detailed couplings were controlled by the α-states. In other words, the original theory with specified couplings and spacetime wormholes was equivalent (from the asymptotic point of view) to a theory with an ensemble of bulk couplings but where spacetime wormholes were forbidden. The same construction will apply in our context, but with one important distinction. Namely, [55–57] focussed on wormholes with Planck-sized cross-sections under the assumption that microscopic wormholes would dominate in any physical process. But the mouths of the replica wormholes in [48, 49] are determined by the location of a quantum extremal surface. As a result, they approximately coincide with the relevant black hole horizons and thus are macroscopic in size. Integrating out such wormholes thus induces an ensemble of highly non-local couplings in the effective action. Indeed, the couplings naturally mediate transitions in which any given interior configuration specifying the geometry and matter fields arbitrarily far inside the black hole can be replaced by any

other, no matter deep the black holes throat may have become. At least for replica numbers n near 1, the action for a replica wormhole whose mouth has area A is of order $\frac{A}{4G}$ [112], so the amplitude for such processes should be exponentially small in this quantity. However, in an old black hole the large number of internal states can lead to a large effect as seen directly above and in [49] (and as foreshadowed in [113–115]).

The exact location and nature of the above non-local interactions is clearly of some interest. In particular, while quantum extremal surfaces may appear outside the black hole's event horizon [45], for black holes evaporating into a vacuum they should always lie inside [42, 43]. Were all of the physics determined by replica wormholes confined far enough inside the horizon, there would be no possibility of affecting the exterior, and in particular no way it could purify the emitted Hawking radiation. However, any separation of the QES from the horizon arises from time dependence, which is typically associated with quantum effects. The backreaction of such effects on the spacetime is then suppressed by a power of G. As a result, the QES tends to be adiabatically close to any horizon, and thus separated by an amount only of order G. In addition, since the QES is determined by balancing the quantum effect of evaporating against a classical effect, the saddle-point is somewhat broad. A rough estimate of the width of the saddle-point suggests that the typical fluctuations of the area are also of order G.[22] This places the QES outside the horizon with order one amplitude. The associated non-local interactions will then naturally transfer information from the deep black hole interior into the outgoing Hawking radiation in much the form suggested in [22, 30].

However, for a full understanding of the physics associated with such interactions it appears one must take into account the corrections they imply for the theory's physical inner product. As described in section 4, such corrections are associated with extending the familiar diffeomorphism invariance of gravitational systems to a more general slicing invariance of the path integral with topology change. Extending this to arbitrary Euclidean time evolution — even involving processes that change the topology of the slice used to define the quantum state — implies spacetimes of different topologies to be gauge related. In other words, this is a restatement of the old maxim that for gravitational systems time evolution is a gauge symmetry unless it involves evolution along an asymptotic boundary. This then directly implies that the path integral com-

[22] For example, we can perform the path integral over replicated geometries and matter, while leaving unfixed the location of the QES where branching between replicas occurs. This leaves a final integral over the QES location to compute, which is roughly $\int e^{-S_{\rm gen}}$ for n close to 1, where $S_{\rm gen}$ is the generalised entropy of the QES and we integrate over its location. The integral over the area of the QES (fixing ingoing time, for example) is then $\int dA\, e^{-S_{\rm gen}(A)}$, with $S_{\rm gen}(A) \sim \frac{A}{4G} + \#\log(A_0 - A)$ [42, 43], where A_0 is the area of the (stretched) horizon. At the saddle point, where $A_0 - A$ is of order G, we have $S''_{\rm gen}(A)$ of order G^{-2} leading to a width ΔA of order G.

putes the gauge invariant physical product as one would expect from general arguments [69, 76–78] (though admittedly those arguments are most direct in contexts where it is not obvious that topology change should be included).

As a result, one may think of the induced nonlocal interactions as modifying the gravitational constraints; i.e., with new terms in the Wheeler-DeWitt equation. The interesting feature, however, is that these modifications are highly non-generic. In the regime that in our models corresponds to $k \gg Z_\alpha$, there are a large number of strongly correlated small corrections, where the correlations conspire to give a large number of null states; i.e., they make the physical inner product highly degenerate so that a priori independent states are in fact linearly dependent in the physical Hilbert space, and so that the dimension of the physical Hilbert space is bounded by Z_α. Furthermore, following ideas related to [96], we argued in section 4 that null states must enforce a similar bound in a general reflection positive gravitational path integral.

It is this bound that leads to the Page curve, and which thus determines the rate at which the above interactions transfer information out of the black hole. As a result, while the above non-local interactions are intimately tied to this change in the inner product, it is natural to think of the former as secondary and the latter as primary. In particular, it is in terms of the inner product that (for reflection positive path integrals) we find a clean statement of the correlations and conspiracies inherent in the details of the induced interactions; see again section 4.

We believe the explicit demonstration of such a large number of null states to be a lesson of fundamental importance. It implies that — due to the above mentioned conspiracies — the gauge symmetry of gravitational systems is much larger and more powerful than had been previously established. The idea that bounds on entropy might be related to such a gauge symmetry date back at least to the early 1990's, when such suggestions arose in discussions of black hole complementarity proposals (see e.g. comments in [104]) and cosmological analogues in de Sitter space. It is also much like the truncation of the bulk Hilbert space implicit in random tensor network models [116, 117] in which the disorder is implemented by inserting randomly chosen projections into the bulk. However, we now see this to be a direct result of the gravitational path integral.

The physics of this enlarged gravitational gauge invariance remains to be understood in detail, especially in the context of more realistic models. Nevertheless, the argument of section 4 indicates that the long discussed relation [11, 96, 118, 119] between two-sided bulk black holes and bulk thermofield double states (4.15) should be understood as an example of this gauge equivalence. In particular, we now see that the so-called "superselection sectors" of [120] — which were argued there to be physically

distinct — are in fact gauge equivalent.[23]

We now speculate further on the implications of this enhanced gauge invariance for issues involving black hole information and the connection to other works. It seems clear that in sufficiently old black holes (where the number of a priori independent internal states is sufficiently large), this gauge invariance implies that vast numbers of a priori independent states must in fact to be regarded as physically equivalent. Furthermore, at least in our model, this happens in an essentially random way that does not respect any additional structure[24]. Extrapolating this result to more complicated models suggests that one will find many states which a priori seem to have very different physics — and in particular in which infalling observers have vastly different experiences — but which are nevertheless gauge equivalent. For example, just as there can be gauge equivalence between Alice meeting Bob and Alice finding only empty space, there is no reason for the physical inner product to respect Alice's notion of particle number (as distinguished, say, from total charges coupled to a gauge field), or even her notion of particle number in a given mode. As a result, even for pure state black holes, the experience of observers inside the black hole may fundamentally fail to be well-defined as a gauge invariant concept. One may view this as a variant of the firewall-like possibility described in [41] that black holes may have 'no interior', or at least no interior from which familiar physics can be extracted.

Nevertheless, as with any gauge symmetry, one is free to fix a gauge in order to define a language (i.e., a set of observables) with which to describe the physics. In particular, as noted above, at the level of Hilbert spaces any gauge invariance is naturally associated with what one may roughly call a projection P from some kinematic Hilbert space $\mathcal{H}_{\rm kin}$ to a physical Hilbert space[25] $\mathcal{H}_{\rm phys} \subset \mathcal{H}_{\rm kin}$. In this sense, one may think of a general gauge fixing procedure as a choice of linear subspace $\mathcal{H}_{\rm GF} \subset \mathcal{H}_{\rm kin}$ such that P defines a bijection between $\mathcal{H}_{\rm GF}$ and $\mathcal{H}_{\rm phys}$. Within a given such gauge fixing scheme, it may then be that the experiences of infalling observers become well-defined. For example, in describing the interior of a black hole of radius R_0 that recently formed from collapse, it would be natural to choose a gauge in which the interior is of size comparable to R_0 (even if such small interiors are gauge equivalent to certain much larger interiors that might form when an initially much larger black hole decays to size R_0), and in particular in which standard effective field theory is a good approximation.

[23] This gauge equivalence resolves a problem noted in that work concerning how such superselection sectors in transform under permutations.

[24] In particular, the spectrum of possibilities allows *any* Hermitian inner product of the appropriate rank.

[25] A structure of this general sort is inherent in Dirac's constraint quantization of gauge systems [70], though the interested reader can consult [71–75] for a variety of more technical treatments.

With this in mind, we recall that the discussions of [42, 44–49, 121] described a close parallel between old black holes that have been radiating into an external system ('the bath') and the ER=EPR paradigm of [31]. In particular, these works suggested that infalling observers experience only standard physics even at the horizon of black holes that have been evaporating for longer than the Page time. At first sight such statements may seem to be in great tension with our bound on the number of linearly independent states inside the black hole. But this tension can be resolved by interpreting the comments of [42, 44–49, 121] as providing a gauge fixed description, where in this case the choice of gauge depends on the state of the bath. In other words, if the black hole system with physical Hilbert space $\mathcal{H}_{\rm phys}$ is considered in the presence of another system with Hilbert space $\mathcal{H}_{\rm bath}$ then, even if the bath system by itself has no gauge invariance, one is free to gauge fix by choosing a general linear subspace $\mathcal{H}_{\rm GF,\,joint} \subset \mathcal{H}_{\rm kin} \otimes \mathcal{H}_{\rm bath}$ for which P defines a bijection to $\mathcal{H}_{\rm phys} \otimes \mathcal{H}_{\rm bath}$. Note that there is no requirement for $\mathcal{H}_{\rm GF,\,joint}$ be a tensor product $\mathcal{H}_{\rm GF_0} \otimes \mathcal{H}_{\rm bath}$ for any fixed subspace $\mathcal{H}_{\rm GF_0} \subset \mathcal{H}_{\rm kin}$. Instead, one is free to effectively let the choice of subspace $\mathcal{H}_{\rm GF_0} \subset \mathcal{H}_{\rm kin}$ vary with the choice of state in $\mathcal{H}_{\rm bath}$.

The connection with the above works is particularly clear in the discussion of Petz reconstruction in [49]. There one wishes to reconstruct an operator $\mathcal{O}$ on $\mathcal{H}_{\rm kin}$ using an operator $\mathcal{O}_R$ on $\mathcal{H}_{\rm bath}$. Now, since $\mathcal{O}_R$ is an operator on $\mathcal{H}_{\rm bath}$, it is automatically gauge invariant. However, since the operators $\mathcal{O}$ discussed in that work were constructed without regard to the (random) physical inner product, they are not gauge invariant. This is consistent, as $\mathcal{O}_R$ reconstructs $\mathcal{O}$ only on a subspace $\mathcal{H}_{\rm code} \subset \mathcal{H}_{\rm kin} \otimes \mathcal{H}_{\rm bath}$ that similarly fails to be gauge invariant. However, at least to good approximation we can think of $\mathcal{H}_{\rm code}$ as defining a partial gauge fixing (meaning that we could choose some $\mathcal{H}_{\rm GF,joint} \supset \mathcal{H}_{\rm code}$. In particular, we may use any bath bra-state $\langle \psi_{\rm bath}|$ to define a linear map from $\mathcal{H}_{\rm code}$ to $\mathcal{H}_{\rm kin}$ via its natural action on $\mathcal{H}_{\rm bath}$. And for any choice of $\langle \psi_{\rm bath}|$, the image defines a subspace $\mathcal{H}_\psi \subset \mathcal{H}_{\rm kin}$ with at most dimension $d_{\rm code} \ll e^{S_{\rm BH}}$, i.e., where this dimension is much less than the dimension of $\mathcal{H}_{\rm phys}$. As a result, with high probability distinct states in $\mathcal{H}_\psi$ will project to distinct states of $\mathcal{H}_{\rm phys}$. In this sense $\mathcal{H}_{\rm code}$ approximately satisfies the requirements for a partial gauge fixing; a complete gauge fixing would result from extending $\mathcal{H}_{\rm code}$ to make the projection of each $\mathcal{H}_\psi$ isomorphic to $\mathcal{H}_{\rm phys}$.

We note that such a gauge fixed interpretation allows all of the hallmarks of what is often called state dependence [25–28] and which is naturally associated with the ER=EPR paradigm. In particular, in contexts where one expects to find only a small number of black hole states (states in $\mathcal{H}_{\rm phys}$) for each bath state, it will be possible to choose a partial gauge fixing of the form described above that selects only states in $\mathcal{H}_{\rm kin}$ with no drama at the horizon. In particular, one will be able to choose a code

subspace within which the evolution can be well-described by standard local effective field theory. In addition, we note that standard objections [40, 41, 122–124] to state dependence focus on non-uniqueness of the predicted physics, and that such objections are clearly moot in a context where the state dependence is simply a choice of gauge (so that non uniqueness of $\mathcal{H}_{\rm GF}$ is to be expected, and so that the gauge invariant predictions are in fact identical).

Nevertheless, the non-uniqueness arguments of [40, 41, 122–124] then show the sort of states that, while they appear at first sight to be physically distinct, must in fact be related by the enlarged gauge symmetry described above. In particular, tracing through such leads to other gauges in which infalling observers experience varying amounts and types of drama at the horizon, as well as to gauges where the observer simply fails to exist in the interior of the black hole.[26] Furthermore, just as there is a particular gauge (or class of gauges) realizing ER=EPR-like scenarios, it seems likely that one can also find gauges realizing fuzzball scenarios (see e.g. [17, 19, 24, 29, 125–129], the non-violent non-locality proposal[27] [20, 22, 30]), proposals emphasizing the bulk Wheeler-DeWitt equation [130, 131], the black hole final state proposal [14], and perhaps other proposals as well.

On the other hand, the above discussion immediately raises the question of how different experiences of a given observer could possibly be gauge related, and thus how the above scenario could possibly be realized in models that are sufficiently realistic to describe our own universe. While there is surely more to be said about this issue, we note that any gauge fixing scheme can be used to *define* an associated gauge invariant observable. I.e., just as one can use Coulomb gauge in electromagnetism to define gauge invariant operators ("the potential in Coulomb gauge"), in the above scenario one can use any gauge to define a notion of observer inside the black hole. The variety of possible gauges would then mean that there are a variety of possible gauge invariant definitions of the observer which happen to coincide (or nearly coincide) under familiar conditions outside old black holes but which differ greatly inside old black holes. One may then rephrase the above statement in a less surprising manner: While we may well-enough understand how to define an observer at the leading semi-classical level,

[26] If one imposes the constraint that the observer survives (in a recognizable form) for a given proper time behind the black hole horizon, then one would expect a generic gauge consistent with this constraint to predict the maximum amount of such drama consistent with the observer's survival to that point.

[27] The non-locality scale L_d in spacetime dimension d is set by the condition $\Delta A \sim G$ described in footnote 22. On a Killing slice of a static black hole of area-radius R, the corresponding proper distance from the event horizon would be $L_d \sim \left(\frac{\ell_p}{R}\right)^{\frac{d-4}{2}} \ell_p$. With respect to the definitions of [22], L_d then gives "non-violent" physics for $d < 4$.

there may be many possible extensions of this definition at the level of non-perturbative physics, and predictions for the observer inside old black holes may depend sensitively on the choice of this extension[28]. The scenario described above (in which apparently distinct observer experiences are gauge related) may thus be considered to be just another version of this idea. It will likely be of great interest to further explore such conjectures and related physics in future work.

Acknowledgments This work was motivated and facilitated by three specific conversations, first with Geoffrey Pennington, second with Xi Dong, and third Steve Giddings, as well as by a long history of discussing black hole information with the entire UCSB High Energy and Gravity group. We also acknowledge interesting conversations with Daniel Harlow, Gary Horowitz, Ted Jacobson, Javier Magán, Juan Maldacena, Xiaoliang Qi, Steve Shenker, Mark Srednicki, Douglas Stanford, Herman Verlinde and Edward Witten. We are grateful for support from NSF grant PHY1801805 and funds from the University of California. H.M. was also supported in part by a DeBenedictis Postdoctoral Fellowship, and D.M. thanks UCSB's KITP for their hospitality during the final portions of this work. As a result, this research was also supported in part by the National Science Foundation under Grant No. NSF PHY-1748958 to the KITP.

A Limits of moments of the Poisson distribution

In this appendix, we study the moments $\langle Z^n \rangle$ of a Poisson random variable Z with mean λ in various limits. This is useful to ascertain the convergence properties of sums $\sum_n c_n |Z^n\rangle$ constructing states of $\mathcal{H}_{\rm BU}$ in section 3, and to illustrate the failure of the third quantised perturbation theory of section 5 in our model.

The moments are given by the Bell polynomials,

$$\mathfrak{Z}^{-1}\langle Z^n \rangle = B_n(\lambda), \tag{A.1}$$

defined by

$$B_n(\lambda) = e^{-\lambda} \sum_{d=0}^{\infty} \frac{\lambda^d}{d!} d^n \,. \tag{A.2}$$

[28]Note that if there is a priori no mechanism for selecting one such definition as preferred, then it is natural to adopt a Bayesian approach and declare that all such extensions are realized with equal probability (or more generally that they are realized according to some probability measure describing the priors of the given theorist studying the system). The question of 'what does an observer experience when falling into a black hole' would then be an inherently probabilistic one, somewhat akin to asking 'what does an observer experience when they are decohered into many Everett branches of the wavefunction of the universe?' We have already conjectured above that with high probability the observer simply fails to exist inside the black hole in a generic gauge, and that post-selecting only on existence of the observer would lead to high drama.

From this, one can check the recurrence relation

$$B_{n+1}(\lambda) = \lambda(B_n'(\lambda) + B_n(\lambda)) \tag{A.3}$$

and $B_0(\lambda) = 1$, from which we can see that $B_n(\lambda)$ is a monic polynomial of order n. In particular this gives us the scaling at large λ and fixed n,

$$B_n(\lambda) \sim \lambda^n, \qquad \lambda \to \infty, \;\; n \text{ fixed}. \tag{A.4}$$

A.1 Large n and convergence

For studying convergence of $\sum_n c_n|Z^n\rangle$, we require the moments at large n and fixed λ. For this, observe that the ratio of consecutive terms in the sum defining $B_n(\lambda)$ is

$$\frac{\lambda}{d}\left(\frac{d}{d-1}\right)^n \sim \frac{\lambda}{d}e^{n/d}, \tag{A.5}$$

where the asymptotic form applies for $1 \ll n \ll d^2$. For large n, the ratio is unity and hence the dth term in the sum is maximal when $d \sim \frac{n}{\log n}$. Substituting this value back into the sum, we can find an estimate of $B_n(\lambda)$ at large n, which we can write as

$$\frac{B_n(\lambda)}{n!} \sim e^{-n\log\log n + o(n)}, \quad n \to \infty, \quad \lambda \text{ fixed}. \tag{A.6}$$

$$\log B_n(\lambda) \sim n\log n - n\log\log n - n + o(n), \quad n \to \infty, \quad \lambda \text{ fixed}. \tag{A.7}$$

For a more carful derivation and many more terms in the expansion, it is convenient to write $d = \frac{n}{\log n}\left(1 + \frac{x}{\sqrt{n}}\right)$ and take the limit of the terms in the sum as $n \to \infty$ at fixed x. In this limit, the series becomes a Gaussian integral in x. From this, we can estimate the norm of the basis state $\||Z^n\rangle\| = \sqrt{\langle Z^n|Z^n\rangle} = e^{-\lambda/2}\sqrt{B_{2n}(\lambda)}$:

$$\log \||Z^n\rangle\| = n\log n - n\log\log n - n(1-\log 2) + o(n) \qquad \text{as } n \to \infty. \tag{A.8}$$

Now we can begin to characterise convergence of sums $\sum c_n|Z^n\rangle$ in the baby universe Hilbert space of section 3.3. By definition, the series converges if the partial sums form a Cauchy sequence. That is,

$$\sum_{n=0}^{\infty} c_n|Z^n\rangle \quad \text{converges} \iff \left\|\sum_{n=n_1}^{n_2} c_n|Z^n\rangle\right\| \to 0 \text{ as } n_1, n_2 \to \infty, \tag{A.9}$$

where in this limit we can take n_1, n_2 to infinity separately at different rates.[29] We will not characterise such series completely, but find a sufficient condition to give us a class of convergent series, and a necessary condition to constrain them.

[29] It may not be that every element of the completion can be represented by such a Cauchy sequence of partial sums. It is false for the 'free' version where we allow only discs and cylinders, replacing the Poisson distribution by its Gaussian approximation: in that case, this class of Cauchy sequences yields only analytic wavefunctions.

First, a necessary condition for convergence (coming from $n_1 = n_2$) is that the norm of individual terms go to zero

$$\text{Convergence} \implies |c_n| \big\| |Z^n\rangle \big\| \to 0 \text{ as } n \to \infty. \tag{A.10}$$

Now, from (A.8), we see that $\big\| |Z^n\rangle \big\|$ is eventually larger than R^n for any $R > 0$, so $|c_n| R^n$ is bounded, which implies that $f(z) := c_n z^n$ converges in the disc $|z| < R$. Since this holds for all R, we find that our series defines an entire analytic function,

$$\sum_{n=0}^{\infty} c_n |Z^n\rangle \quad \text{converges} \implies f(z) = \sum c_n z^n \text{ is entire analytic.} \tag{A.11}$$

We can thus characterise convergent series in terms of the class of allowed analytic functions. Improving on the analyticity result, we can bound the growth of allowed functions f. To do this, we introduce the order of an analytic function, which is the infimum over all ρ such that $|f(z)| < \exp(|z|^\rho)$ for sufficiently large z. We can strengthen our necessary condition to

$$\sum_{n=0}^{\infty} c_n |Z^n\rangle \quad \text{converges} \implies f(z) = \sum c_n z^n \text{ has order} \leq 1, \tag{A.12}$$

which means that for every $\epsilon > 0$, we have $|f(z)| < \exp(|z|^{1+\epsilon})$ for sufficiently large $|z|$. To show this, we use a result expressing the order in terms of the Taylor coefficients, namely $\text{order}(f) = \limsup_{n\to\infty} \frac{n \log n}{\log(1/|c_n|)}$. For the norm of the terms in the series to go to zero, we must have $\log(1/|c_n|) - \log \big\| |Z^n\rangle \big\|$ go to infinity, so for sufficiently large n we have $\log(1/|c_n|) > \log \big\| |Z^n\rangle \big\|$. From (A.8), for any $\epsilon > 0$ and sufficiently large n we have $\log \big\| |Z^n\rangle \big\| > (1-\epsilon) n \log n$. In turn, this means that $\log(1/|c_n|) > (1-\epsilon) n \log n$ for large enough n, and hence $\limsup_{n\to\infty} \frac{n \log n}{\log(1/|c_n|)} \leq 1$.

Our sufficient condition is absolute convergence, which means that the sum of norms converges, and follows from the triangle inequality for the norm.

$$\sum_n |c_n| \big\| |Z^n\rangle \big\| \text{ convergent} \implies \sum_n c_n |Z^n\rangle \text{ convergent.} \tag{A.13}$$

Now, from (A.8), we have the result that $\big\| |Z^n\rangle \big\|$ decays faster than $n! a^n$ for any a. From this, we can find a simple sufficient bound on the coefficients for convergence,

$$|c_n| < A \frac{x^n}{n!} \text{ for some } A, x \implies \sum_n c_n |Z^n\rangle \text{ convergent.} \tag{A.14}$$

In particular, this means that any exponential function $|e^{xZ}\rangle$, or more generally a function of exponential type, defines a convergent series by its Taylor expansion.

The gap between our sufficient and necessary conditions (order one functions that are not of exponential type) is small but nonempty, for example containing $\frac{1}{\Gamma(-z)}$.

A.2 Large λ and n

Here, we study a limit of $\lambda \to \infty$ and $n \to \infty$ at fixed ratio $\nu = \frac{n}{\lambda}$, which will interpolate between the large λ fixed n and large n fixed λ results. We could proceed from the same series expression, but we use an alternative method, starting from an integral representation of $B_n(\lambda)$. This expression extracts the moments from the generating function (3.11) by a contour integral

$$\frac{B_n(\lambda)}{n!} = \frac{1}{2\pi i} \oint \frac{du}{u^{n+1}} e^{\lambda(e^u - 1)}, \tag{A.15}$$

where the contour encircles the origin. We can evaluate this by steepest descent, looking for stationary points of

$$S(u) = e^u - 1 - \nu \log u \,. \tag{A.16}$$

The stationary points $S'(u) = 0$ solve $\nu = ue^u$, and the relevant saddle point is the unique positive solution, which defines the Lambert W function or product logarithm,

$$u_* = W(\nu). \tag{A.17}$$

Applying the steepest descent method at this saddle point gives us

$$\frac{B_n(\lambda)}{n!} \sim \frac{e^{\lambda S(u_*)}}{u_* \sqrt{2\pi S''(u_*)\lambda}} \,. \tag{A.18}$$

This result in fact interpolates between our two previous results for large λ fixed n (by taking $\nu \ll 1$) and large n fixed λ (by taking $\nu \gg 1$).

It is interesting in particular to see how the large λ result breaks down when n becomes large. Taking $\nu \ll 1$ we have $u_* = \nu - \nu^2 + O(\nu^3)$, so $S(u_*) \sim -\nu \log \nu + \nu + \frac{1}{2}\nu^2 + \cdots$, with higher terms all integer powers of ν. Substituting this into the steepest descent result, we have

$$\frac{e^{\lambda S(u_*)}}{u_* \sqrt{2\pi S''(u_*)\lambda}} \sim \frac{e^{n \log \lambda - n \log n + n + \frac{n^2}{2\lambda} + \cdots}}{\sqrt{2\pi n}} \sim \frac{\lambda^n}{n!} e^{\frac{n^2}{2\lambda} + \cdots}, \tag{A.19}$$

where we applied Stirling's approximation to the factorial. The terms in the exponential are of the form $\frac{n^k}{\lambda^{k-1}}$ for $k = 2, 3, \ldots$, and become relevant when n is of order $\lambda^{1-1/k}$. The first correction occurs from the $k = 2$ term shown explicitly, first relevant when n is of order $\sqrt{\lambda}$, when it contributes an order one rescaling of $B_n(\lambda)$:

$$B_n(\lambda) \sim \lambda^n e^{\frac{n^2}{2\lambda}}, \qquad \lambda \to \infty, \frac{n^2}{\lambda} \text{ fixed}. \tag{A.20}$$

Higher order terms in the exponential are given by higher orders in the expansion of $S(u_*)$ at small ν.

References

[1] D. N. Page, *Average entropy of a subsystem*, *Phys. Rev. Lett.* **71** (1993) 1291–1294, [gr-qc/9305007].

[2] T. Jacobson, *Introduction to quantum fields in curved space-time and the Hawking effect*, in *Lectures on quantum gravity. Proceedings, School of Quantum Gravity, Valdivia, Chile, January 4-14, 2002*, pp. 39–89, 2003. gr-qc/0308048. DOI.

[3] S. D. Mathur, *The Information paradox: A Pedagogical introduction*, *Class. Quant. Grav.* **26** (2009) 224001, [0909.1038].

[4] D. Harlow, *Jerusalem Lectures on Black Holes and Quantum Information*, *Rev. Mod. Phys.* **88** (2016) 015002, [1409.1231].

[5] W. G. Unruh and R. M. Wald, *Information Loss*, *Rept. Prog. Phys.* **80** (2017) 092002, [1703.02140].

[6] D. Marolf, *The Black Hole information problem: past, present, and future*, *Rept. Prog. Phys.* **80** (2017) 092001, [1703.02143].

[7] L. Susskind, L. Thorlacius and J. Uglum, *The Stretched horizon and black hole complementarity*, *Phys. Rev.* **D48** (1993) 3743–3761, [hep-th/9306069].

[8] L. Susskind, *String theory and the principles of black hole complementarity*, *Phys. Rev. Lett.* **71** (1993) 2367–2368, [hep-th/9307168].

[9] L. Susskind, *Strings, black holes and Lorentz contraction*, *Phys. Rev.* **D49** (1994) 6606–6611, [hep-th/9308139].

[10] G. Chapline, E. Hohlfeld, R. B. Laughlin and D. I. Santiago, *Quantum phase transitions and the breakdown of classical general relativity*, *Int. J. Mod. Phys.* **A18** (2003) 3587–3590, [gr-qc/0012094].

[11] J. M. Maldacena, *Eternal black holes in anti-de Sitter*, *JHEP* **04** (2003) 021, [hep-th/0106112].

[12] P. O. Mazur and E. Mottola, *Gravitational condensate stars: An alternative to black holes*, gr-qc/0109035.

[13] F. Winterberg, *Gamma Ray Bursters and Lorentzian Relativity*, *Z. Naturforsch.* **56A** (2001) 889–892.

[14] G. T. Horowitz and J. M. Maldacena, *The Black hole final state*, *JHEP* **02** (2004) 008, [hep-th/0310281].

[15] S. W. Hawking, *Information loss in black holes*, *Phys. Rev.* **D72** (2005) 084013, [hep-th/0507171].

[16] G. T. Horowitz and E. Silverstein, *The Inside story: Quasilocal tachyons and black holes*, *Phys. Rev.* **D73** (2006) 064016, [hep-th/0601032].

[17] S. D. Mathur, *The fuzzball proposal for black holes: An elementary review*, *Fortsch. Phys.* **53** (2005) 793–827, [hep-th/0502050].

[18] S. B. Giddings, *Black hole information, unitarity, and nonlocality*, *Phys. Rev.* **D74** (2006) 106005, [hep-th/0605196].

[19] S. D. Mathur, *Fuzzballs and the information paradox: A Summary and conjectures*, 0810.4525.

[20] S. B. Giddings, *Models for unitary black hole disintegration*, *Phys. Rev.* **D85** (2012) 044038, [1108.2015].

[21] A. Davidson, *Holographic Shell Model: Stack Data Structure inside Black Holes?*, *Int. J. Mod. Phys.* **D23** (2014) 1450041, [1108.2650].

[22] S. B. Giddings, *Nonviolent nonlocality*, *Phys. Rev.* **D88** (2013) 064023, [1211.7070].

[23] A. Almheiri, D. Marolf, J. Polchinski and J. Sully, *Black Holes: Complementarity or Firewalls?*, *JHEP* **02** (2013) 062, [1207.3123].

[24] S. D. Mathur and D. Turton, *Comments on black holes I: The possibility of complementarity*, *JHEP* **01** (2014) 034, [1208.2005].

[25] K. Papadodimas and S. Raju, *An Infalling Observer in AdS/CFT*, *JHEP* **10** (2013) 212, [1211.6767].

[26] E. Verlinde and H. Verlinde, *Black Hole Entanglement and Quantum Error Correction*, *JHEP* **10** (2013) 107, [1211.6913].

[27] Y. Nomura, J. Varela and S. J. Weinberg, *Complementarity Endures: No Firewall for an Infalling Observer*, *JHEP* **03** (2013) 059, [1207.6626].

[28] E. Verlinde and H. Verlinde, *Passing through the Firewall*, 1306.0515.

[29] S. D. Mathur and D. Turton, *The flaw in the firewall argument*, *Nucl. Phys.* **B884** (2014) 566–611, [1306.5488].

[30] S. B. Giddings, *Nonviolent information transfer from black holes: A field theory parametrization*, *Phys. Rev.* **D88** (2013) 024018, [1302.2613].

[31] J. Maldacena and L. Susskind, *Cool horizons for entangled black holes*, *Fortsch. Phys.* **61** (2013) 781–811, [1306.0533].

[32] E. Silverstein, *Backdraft: String Creation in an Old Schwarzschild Black Hole*, 1402.1486.

[33] C. Rovelli and F. Vidotto, *Planck stars*, *Int. J. Mod. Phys.* **D23** (2014) 1442026, [1401.6562].

[34] H. M. Haggard and C. Rovelli, *Quantum-gravity effects outside the horizon spark black to white hole tunneling*, *Phys. Rev.* **D92** (2015) 104020, [1407.0989].

[35] A. Giveon, N. Itzhaki and D. Kutasov, *Stringy Horizons*, *JHEP* **06** (2015) 064, [1502.03633].

[36] S. W. Hawking, M. J. Perry and A. Strominger, *Soft Hair on Black Holes*, *Phys. Rev. Lett.* **116** (2016) 231301, [1601.00921].

[37] M. Christodoulou, C. Rovelli, S. Speziale and I. Vilensky, *Planck star tunneling time: An astrophysically relevant observable from background-free quantum gravity*, *Phys. Rev.* **D94** (2016) 084035, [1605.05268].

[38] A. Giveon and N. Itzhaki, *Stringy Information and Black Holes*, 1912.06538.

[39] L. Amadei and A. Perez, *Hawking's information puzzle: a solution realized in loop quantum cosmology*, 1911.00306.

[40] A. Almheiri, D. Marolf, J. Polchinski, D. Stanford and J. Sully, *An Apologia for Firewalls*, *JHEP* **09** (2013) 018, [1304.6483].

[41] D. Marolf and J. Polchinski, *Gauge/Gravity Duality and the Black Hole Interior*, *Phys. Rev. Lett.* **111** (2013) 171301, [1307.4706].

[42] G. Penington, *Entanglement Wedge Reconstruction and the Information Paradox*, 1905.08255.

[43] 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, [1905.08762].

[44] A. Almheiri, R. Mahajan, J. Maldacena and Y. Zhao, *The Page curve of Hawking radiation from semiclassical geometry*, 1908.10996.

[45] A. Almheiri, R. Mahajan and J. Maldacena, *Islands outside the horizon*, 1910.11077.

[46] A. Almheiri, R. Mahajan and J. E. Santos, *Entanglement islands in higher dimensions*, 1911.09666.

[47] H. Z. Chen, Z. Fisher, J. Hernandez, R. C. Myers and S.-M. Ruan, *Information Flow in Black Hole Evaporation*, 1911.03402.

[48] A. Almheiri, T. Hartman, J. Maldacena, E. Shaghoulian and A. Tajdini, *Replica Wormholes and the Entropy of Hawking Radiation*, 1911.12333.

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

[50] S. Ryu and T. Takayanagi, *Aspects of Holographic Entanglement Entropy*, *JHEP* **08** (2006) 045, [hep-th/0605073].

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

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

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

[54] X. Dong and A. Lewkowycz, *Entropy, Extremality, Euclidean Variations, and the Equations of Motion*, *JHEP* **01** (2018) 081, [1705.08453].

[55] S. R. Coleman, *Black Holes as Red Herrings: Topological Fluctuations and the Loss of Quantum Coherence*, *Nucl. Phys.* **B307** (1988) 867–882.

[56] S. B. Giddings and A. Strominger, *Loss of Incoherence and Determination of Coupling Constants in Quantum Gravity*, *Nucl. Phys.* **B307** (1988) 854–866.

[57] S. B. Giddings and A. Strominger, *Baby Universes, Third Quantization and the Cosmological Constant*, *Nucl. Phys.* **B321** (1989) 481–508.

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

[59] S. S. Gubser, I. R. Klebanov and A. M. Polyakov, *Gauge theory correlators from noncritical string theory*, *Phys. Lett.* **B428** (1998) 105–114, [hep-th/9802109].

[60] E. Witten, *Anti-de Sitter space and holography*, *Adv. Theor. Math. Phys.* **2** (1998) 253–291, [hep-th/9802150].

[61] G. W. Gibbons and S. W. Hawking, *Action Integrals and Partition Functions in Quantum Gravity*, *Phys. Rev.* **D15** (1977) 2752–2756.

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

[63] N. Arkani-Hamed, J. Orgera and J. Polchinski, *Euclidean wormholes in string theory*, *JHEP* **12** (2007) 018, [0705.2768].

[64] S. W. Hawking, *Quantum Coherence Down the Wormhole*, *Phys. Lett.* **B195** (1987) 337.

[65] S. W. Hawking, *Wormholes in Space-Time*, *Phys. Rev.* **D37** (1988) 904–910.

[66] S. B. Giddings and A. Strominger, *Axion Induced Topology Change in Quantum Gravity and String Theory*, *Nucl. Phys.* **B306** (1988) 890–907.

[67] G. V. Lavrelashvili, V. A. Rubakov and P. G. Tinyakov, *Disruption of Quantum Coherence upon a Change in Spatial Topology in Quantum Gravity*, *JETP Lett.* **46** (1987) 167–169.

[68] J. B. Hartle and S. W. Hawking, *Wave Function of the Universe*, *Phys. Rev.* **D28** (1983) 2960–2975.

[69] J. J. Halliwell and J. B. Hartle, *Wave functions constructed from an invariant sum over histories satisfy constraints*, *Phys. Rev.* **D43** (1991) 1170–1194.

[70] P. A. M. Dirac, *Lectures on Quantum Mechanics*. Belfor Graduate School of Science, Yeshiva University, New York, 1964.

[71] N. P. Landsman, *Rieffel induction as generalized quantum Marsden-Weinstein reduction*, hep-th/9305088.

[72] D. Marolf, *Quantum observables and recollapsing dynamics*, *Class. Quant. Grav.* **12** (1995) 1199–1220, [gr-qc/9404053].

[73] A. Ashtekar, J. Lewandowski, D. Marolf, J. Mourao and T. Thiemann, *Quantization of diffeomorphism invariant theories of connections with local degrees of freedom*, *J. Math. Phys.* **36** (1995) 6456–6493, [gr-qc/9504018].

[74] D. Marolf, *Group averaging and refined algebraic quantization: Where are we now?*, in *Recent developments in theoretical and experimental general relativity, gravitation and relativistic field theories. Proceedings, 9th Marcel Grossmann Meeting, MG'9, Rome, Italy, July 2-8, 2000. Pts. A-C*, 2000. gr-qc/0011112.

[75] O. Yu. Shvedov, *On correspondence of BRST-BFV, Dirac and refined algebraic quantizations of constrained systems*, *Annals Phys.* **302** (2002) 2–21, [hep-th/0111270].

[76] D. Marolf, *Path integrals and instantons in quantum gravity: Minisuperspace models*, *Phys. Rev.* **D53** (1996) 6979–6990, [gr-qc/9602019].

[77] M. P. Reisenberger and C. Rovelli, *'Sum over surfaces' form of loop quantum gravity*, *Phys. Rev.* **D56** (1997) 3490–3508, [gr-qc/9612035].

[78] J. B. Hartle and D. Marolf, *Comparing formulations of generalized quantum mechanics for reparametrization - invariant systems*, *Phys. Rev.* **D56** (1997) 6247–6257, [gr-qc/9703021].

[79] A. S. Wightman, *Quantum Field Theory in Terms of Vacuum Expectation Values*, *Phys. Rev.* **101** (1956) 860–866.

[80] R. F. Streater and A. S. Wightman, *PCT, spin and statistics, and all that*. Princeton University Press, 2016.

[81] K. Osterwalder and R. Schrader, *AXIOMS FOR EUCLIDEAN GREEN'S FUNCTIONS*, *Commun. Math. Phys.* **31** (1973) 83–112.

[82] A. Blommaert, T. G. Mertens and H. Verschelde, *Eigenbranes in Jackiw-Teitelboim gravity*, 1911.11603.

[83] D. Harlow, *Wormholes, Emergent Gauge Fields, and the Weak Gravity Conjecture*, *JHEP* **01** (2016) 122, [1510.07911].

[84] M. Guica and D. L. Jafferis, *On the construction of charged operators inside an eternal black hole*, *SciPost Phys.* **3** (2017) 016, [1511.05627].

[85] D. Harlow and D. Jafferis, *The Factorization Problem in Jackiw-Teitelboim Gravity*, 1804.01081.

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

[87] D. Stanford and E. Witten, *JT Gravity and the Ensembles of Random Matrix Theory*, 1907.03363.

[88] R. Dijkgraaf, H. L. Verlinde and E. P. Verlinde, *Notes on topological string theory and 2-D quantum gravity*, in *Cargese Study Institute: Random Surfaces, Quantum Gravity and Strings Cargese, France, May 27-June 2, 1990*, pp. 0091–156, 1990.

[89] A. F. Möbius, *Theorie der elementaren verwandtschaft*, *Berichte über die Verhandlungen der Königlich Sächsischen Gesellschaft der Wissenschaften, Mathematisch-physikalische Klasse* **15** (1863) 19–57.

[90] C. Jordan, *Sur la déformation des surfaces.*, *Journal de mathématiques pures et appliquées* (1866) 105–109.

[91] "Complex wishart distribution." Wikipedia, https://en.wikipedia.org/wiki/Complex_Wishart_distribution.

[92] P. Graczyk, G. Letac, H. Massam et al., *The complex wishart distribution and the symmetric group*, *The Annals of Statistics* **31** (2003) 287–309.

[93] J. Maldacena, D. Stanford and Z. Yang, *Conformal symmetry and its breaking in two dimensional Nearly Anti-de-Sitter space*, *PTEP* **2016** (2016) 12C104, [1606.01857].

[94] J. Engelsy, T. G. Mertens and H. Verlinde, *An investigation of AdS_2 backreaction and holography*, *JHEP* **07** (2016) 139, [1606.03438].

[95] D. Stanford and E. Witten, *Fermionic Localization of the Schwarzian Theory*, *JHEP* **10** (2017) 008, [1703.04612].

[96] D. L. Jafferis, *Bulk reconstruction and the Hartle-Hawking wavefunction*, 1703.01519.

[97] J. Polchinski, *String theory and black hole complementarity*, in *Future perspectives in string theory. Proceedings, Conference, Strings'95, Los Angeles, USA, March 13-18, 1995*, pp. 417–426, 1995. hep-th/9507094.

[98] S. W. Hawking, *Breakdown of Predictability in Gravitational Collapse*, *Phys. Rev.* **D14** (1976) 2460–2473.

[99] J. Polchinski and A. Strominger, *A Possible resolution of the black hole information puzzle*, *Phys. Rev.* **D50** (1994) 7403–7409, [hep-th/9407008].

[100] W. Fischler, I. R. Klebanov, J. Polchinski and L. Susskind, *Quantum Mechanics of the Googolplexus*, *Nucl. Phys.* **B327** (1989) 157–177.

[101] J. Preskill, *Wormholes in Space-time and the Constants of Nature*, *Nucl. Phys.* **B323** (1989) 141–186.

[102] J. S. Cotler, G. Gur-Ari, M. Hanada, J. Polchinski, P. Saad, S. H. Shenker et al., *Black Holes and Random Matrices*, *JHEP* **05** (2017) 118, [1611.04650].

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

[104] D. A. Lowe, J. Polchinski, L. Susskind, L. Thorlacius and J. Uglum, *Black hole complementarity versus locality*, *Phys. Rev.* **D52** (1995) 6997–7010, [hep-th/9506138].

[105] N. Goheer, M. Kleban and L. Susskind, *The Trouble with de Sitter space*, *JHEP* **07** (2003) 056, [hep-th/0212209].

[106] A. Maloney, *Geometric Microstates for the Three Dimensional Black Hole?*, 1508.04079.

[107] A. Almheiri, *Holographic Quantum Error Correction and the Projected Black Hole Interior*, 1810.02055.

[108] Z. Fu and D. Marolf, *Bag-of-gold spacetimes, Euclidean wormholes, and inflation from domain walls in AdS/CFT*, *JHEP* **11** (2019) 040, [1909.02505].

[109] C. Akers, N. Engelhardt, G. Penington and M. Usatyuk, *Quantum Maximin Surfaces*, 1912.02799.

[110] J. Cotler and K. Jensen, *Emergent unitarity in de Sitter from matrix integrals*, 1911.12358.

[111] I. Heemskerk, J. Penedones, J. Polchinski and J. Sully, *Holography from Conformal Field Theory*, *JHEP* **10** (2009) 079, [0907.0151].

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

[113] S. D. Mathur, *A model with no firewall*, 1506.04342.

[114] S. B. Giddings, *Nonviolent unitarization: basic postulates to soft quantum structure of black holes*, *JHEP* **12** (2017) 047, [1701.08765].

[115] S. D. Mathur, *Resolving the black hole causality paradox*, *Gen. Rel. Grav.* **51** (2019) 24, [1703.03042].

[116] P. Hayden, S. Nezami, X.-L. Qi, N. Thomas, M. Walter and Z. Yang, *Holographic duality from random tensor networks*, *JHEP* **11** (2016) 009, [1601.01694].

[117] X.-L. Qi and Z. Yang, *Space-time random tensor networks and holographic duality*, 1801.05289.

[118] M. Van Raamsdonk, *Comments on quantum gravity and entanglement*, 0907.2939.

[119] M. Van Raamsdonk, *Building up spacetime with quantum entanglement*, *Gen. Rel. Grav.* **42** (2010) 2323–2329, [1005.3035].

[120] D. Marolf and A. C. Wall, *Eternal Black Holes and Superselection in AdS/CFT*, *Class. Quant. Grav.* **30** (2013) 025001, [1210.3590].

[121] Y. Chen, *Pulling Out the Island with Modular Flow*, 1912.02210.

[122] R. Bousso, *Firewalls from double purity*, *Phys. Rev.* **D88** (2013) 084035, [1308.2665].

[123] R. Bousso, *Violations of the Equivalence Principle by a Nonlocally Reconstructed Vacuum at the Black Hole Horizon*, *Phys. Rev. Lett.* **112** (2014) 041102, [1308.3697].

[124] D. Marolf and J. Polchinski, *Violations of the Born rule in cool state-dependent horizons*, *JHEP* **01** (2016) 008, [1506.01337].

[125] I. Bena and N. P. Warner, *Black holes, black rings and their microstates*, *Lect. Notes Phys.* **755** (2008) 1–92, [hep-th/0701216].

[126] V. Balasubramanian, J. de Boer, S. El-Showk and I. Messamah, *Black Holes as Effective Geometries*, *Class. Quant. Grav.* **25** (2008) 214004, [0811.0263].

[127] K. Skenderis and M. Taylor, *The fuzzball proposal for black holes*, *Phys. Rept.* **467** (2008) 117–171, [0804.0552].

[128] B. D. Chowdhury and A. Virmani, *Modave Lectures on Fuzzballs and Emission from the D1-D5 System*, 1001.1444.

[129] I. Bena and N. P. Warner, *Resolving the Structure of Black Holes: Philosophizing with a Hammer*, 1311.4538.

[130] T. Jacobson, *Boundary unitarity and the black hole information paradox*, *Int. J. Mod. Phys.* **D22** (2013) 1342002, [1212.6944].

[131] T. Jacobson and P. Nguyen, *Diffeomorphism invariance and the black hole information paradox*, *Phys. Rev.* **D100** (2019) 046002, [1904.04434].