Constraints on Cosmology and Quantum Gravity from Quantum Mechanics and Quantum Field Theory

Thesis by
Jason Pollack

In Partial Fulfillment of the Requirements for the
Degree of
Doctor of Philosophy

Caltech

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2017
Defended May 19, 2017

Acknowledgments

Many individuals and groups of people helped shape my time at Caltech:

- The Friday night dinner/drinks/D&D group: Dustin Anderson, Kevin Barkett, Tony Bartolotta, Jonathan Blackman, Daniel Brooks, and Jeremy Brouillet.

- The Sunday night TV-watching group: Joe Lozier, Dan Pershey, and Paul Plucinsky.

- My parents, for their continuing support of my pursuit of academia, despite the transformation over the last five years of even the titles of my papers into incomprehensibility.

- My fellow Ph 20/21/22 TAs: Casey Handmer, Jenia Mozgunov, and Masha Okounkova; and

- The Ph 20/21/22 instructors, Alan Weinstein and Tom Prince, who were always willing to let us do our own thing and practically run the course, making it both great fun and great experience to TA.

- Senior grad students: Tristan Mckinney, Ingmar Saberi, Chia Hsien Shen, and especially Kim Boddy.

- Junior grad students: Charles Cao, Aidan Chatwin-Davies, and Ashmeet Singh.

- Grad students in my year: Enrico Herrmann, Nick Hunter-Jones, Murat Kologlu, Grant Remmen, and especially my officemate, Tony Bartolotta.

- Postdocs: Stefan Leichenauer in my first years at Caltech and Ning Bao more recently, mentors, inspirations, and dispensers of invaluable advice on how to survive grad school and take the next steps forward into academia.

- Cliff Cheung and Mark Wise, for sharing some of their vast knowledge of quantum field theory with me over the years, and for always having their doors open for me to wander in and ask a question when I was stuck.

he glued himself to the printer and that time he baked his laptop in the oven;

- And, of course, my advisor, Sean Carroll. From the time I wandered into his office before the beginning of my first year and was immediately ushered into a collaboration with him and Stefan, through what must have been dozens of hours of meetings with him and Kim where we hashed out just what the late-time structure of de Sitter space and its derivatives looked like, to his unabating support of my career as I applied for postdocs, I couldn't have asked for a more supportive and collaborative academic partner and mentor. That's not even to mention the many speaker dinners, or Cosmo QM, or his patient toleration of my decision to become GSC chair. As I continue on with my career to UBC and wherever comes after that, Sean will be the model I'll always try to emulate for how to collaborate, how to advise, and how to do physics.

Abstract

Typical cosmological states have structure, obey to very good approximation the laws of classical physics on large scales, and are far from equilibrium. Typical quantum-mechanical states have none of these properties. If the universe is described by a state in a Hilbert space, the state and its Hilbert space must therefore obey a number of constraints to describe realistic cosmological spacetimes. In particular, they must admit a quantum-to-classical transition via decoherence that allows for the emergence of classical spacetimes, and such spacetimes must obey gravitational constraints, in particular on the entanglement entropy of subsystems within them. The papers collected in this thesis are concerned with these constraints. We investigate two holographic correspondences inspired by AdS/CFT, the AdS-MERA correspondence, which suggests that anti-de Sitter space may be given a discretized description as a tensor network, and the ER=EPR duality, which identified entangled qubits with wormholes connecting them. In the former case, we use holographic entropy bounds to severely constrain the properties of any such tensor network; in the latter case we prove a new general-relativistic area theorem which states that an area corresponding to the entanglement entropy in wormhole geometries is exactly conserved. We use information-theoretic constraints to show that under mild assumptions about the black hole interior an observer falling beyond the horizon is unable to verify the claimed cloning of information in the firewall paradox before reaching the singularity. Finally, we analyze the decoherence structures of late-time de Sitter space and early-time slow-roll eternal inflation. We show that in the former case a universe with an infinite-dimensional Hilbert space and a positive cosmological constant inevitably reaches a maximum-entropy state from which no further branching or decoherence is possible, forbidding the existence of dynamical quantum fluctuations at late time. In the latter case, gravitational-strength interaction among inflaton modes leads to decoherence of sufficiently super-Hubble modes, which we argue backreacts to cause different histories of cosmological evolution on different branches and hence creates the conditions necessary for eternal inflation.

Published Content and Contributions

The chapters in this thesis, aside from the Introduction, originally appeared as the following publications and/or arXiv submissions. All authors contributed equally.

- N. Bao, C. Cao, S. M. Carroll, A. Chatwin-Davies, N. Hunter-Jones, J. Pollack and G. N. Remmen, "Consistency conditions for an AdS multiscale entanglement renormalization ansatz correspondence," *Phys. Rev. D* **91**, no. 12, 125036 (2015), `arXiv:1504.06632 [hep-th]`.

- N. Bao, J. Pollack and G. N. Remmen, "Splitting Spacetime and Cloning Qubits: Linking No-Go Theorems across the ER=EPR Duality," *Fortsch. Phys.* **63**, 705 (2015), `arXiv:1506.08203 [hep-th]`.

- N. Bao, J. Pollack and G. N. Remmen, "Wormhole and Entanglement (Non-) Detection in the ER=EPR Correspondence," *JHEP* **1511**, 126 (2015), `arXiv:1509.05426 [hep-th]`.

- G. Remmen, N. Bao and J. Pollack, "Entanglement Conservation, ER=EPR, and a New Classical Area Theorem for Wormholes," *JHEP* **1607**, 048 (2016), `arXiv:1604.08217 [hep-th]`.

- N. Bao, A. Bouland, A. Chatwin-Davies, J. Pollack and H. Yuen, "Rescuing Complementarity With Little Drama," *JHEP* **1612**, 026 (2016), `arXiv:1607.05141 [hep-th]`.

- K. K. Boddy, S. M. Carroll and J. Pollack, "De Sitter Space Without Dynamical Quantum Fluctuations," *Found. Phys.* **46**, no. 6, 702 (2016) `arXiv:1405.0298 [hep-th]`.

- K. K. Boddy, S. M. Carroll and J. Pollack, "How Decoherence Affects the Probability of Slow-Roll Eternal Inflation," *Phys. Rev. D* (submitted), `arXiv:1612.04894 [hep-th]`.

Contents

List of Figures

Introduction

The papers collected together in this thesis have at least two things in common. First, they were written (at least in part) by the author of this thesis. Second, they are concerned (at least in part) with attempts to constrain cosmology through the use of quantum mechanics. The basic idea is to embrace the message from quantum mechanics that Hilbert space is fundamental. Instead of starting from cosmological objects or a spacetime metric, we take the position that such structures must emerge as features inside the wave function of the universe, which is "just" a quantum state in a Hilbert space. But, of course, most quantum states do not look like the current state of our universe as we perceive it! In particular, we observe that the universe has *structure*, that it is (approximately) *classical*, and that it is *far from equilibrium*. Each of these properties place constraints on the quantum-mechanical structure of cosmological systems, and the hope of the broad program of which the papers in this thesis are a part is that collecting and implementing these constraints will allow us to better understand and constrain even highly classical cosmologies.

In quantum-mechanical language, the fact that the universe has structure means that the Hilbert space of the universe does as well: it can be decomposed into a tensor product of very many subsystems, each themselves smaller Hilbert spaces, and the time evolution of states in this Hilbert space (at least the ones that look similar to our own) can be understood as a sum of interactions between only small numbers of these subsystems. As an example, consider quantum field theories. The Hilbert space of a QFT is of course infinite-dimensional, which leads to a number of well-known problems, but more importantly for our purposes it is highly structured. Consider, to specialize further, the Hamiltonian for a free scalar field in four-dimensional Minkowski space:

$$\hat{H} = \int \frac{d^3p}{(2\pi)^3} \omega_{\mathbf{p}} \left(\hat{a}_{\mathbf{p}}^\dagger \hat{a}_{\mathbf{p}} + \frac{1}{2} \left[\hat{a}_{\mathbf{p}}, \hat{a}_{\mathbf{p}}^\dagger \right] \right). \tag{1.1}$$

The Hilbert space of this theory is the product of multiparticle Hilbert spaces at each

point in (Fourier) space; the Hamiltonian[1] is the sum of terms that contain operators $\hat{a}_{\mathbf{p}}$, $\hat{a}_{\mathbf{p}}^{\dagger}$ which act on the individual Hilbert spaces. This theory also has additional symmetries, such as rotation and translation invariance.

When the Hilbert space of a quantum-mechanical theory decomposes, the data associated with a state $|\Psi\rangle$ in this Hilbert space becomes much richer. In particular, when the Hilbert space can be divided into two pieces, $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_{\bar{A}}$ (each of which might be able to be further decomposed), we can construct a reduced density matrix by tracing out one of these pieces:

$$\rho_A \equiv \text{Tr}_{\bar{A}} \, \rho = \text{Tr}_{\bar{A}} \, |\Psi\rangle\langle\Psi|. \tag{1.2}$$

In particular, we can characterize ρ_A in a basis-independent way by computing its *entanglement entropy*,

$$S_A \equiv -\text{Tr}_A \left(\rho_A \ln \rho_A \right). \tag{1.3}$$

The entanglement entropy will be crucial for much of the papers in this thesis, as will the reduced density matrix, especially its evolution with time. These concepts are key for making contact with the classical world, as well as with gravitational theories.

First consider (approximate) classicality. Throughout this thesis we will equate classicality with *decoherence* [3–7]: the dynamical emergence from an overall wave function of multiple distinct noninterfering branches. In the simplest example of decoherence, we partition the Hilbert space into system and environment

$$\mathcal{H} = \mathcal{H}_S \otimes \mathcal{H}_E. \tag{1.4}$$

An arbitrary state $|\Psi\rangle \in \mathcal{H}$ can be written

$$|\Psi\rangle = \sum_{i \leq |\mathcal{H}_S|, j \leq |\mathcal{H}_E|} c_{ij} |S_i\rangle |E_j\rangle, \tag{1.5}$$

where $\{S_i\}$ and $\{E_j\}$ are complete orthonormal bases for $\mathcal{H}_S$ and $\mathcal{H}_E$, respectively. The expression (1.5) is not unique: we can change the coefficients by changing our choices of bases, and in particular linear algebra guarantees that we can always find

---

[1] Even if we didn't know about this decomposition, we might be able to infer it from the spectrum of the Hamiltonian; see [1, 2] for recent progress in this direction.

some special choice of bases (the Schmidt decomposition) for which

$$|\Psi\rangle = \sum_{i \leq \min(|\mathcal{H}_S|,|\mathcal{H}_E|)} c_i |S_i'\rangle |E_i'\rangle. \tag{1.6}$$

Then the branches $|S_i'\rangle|E_i'\rangle$ are orthogonal (and thus noninterfering). However, this is not a dynamical statement: in general, the action of the Hamiltonian will not respect this decomposition, and in particular product states will not evolve into product states. Instead, decoherence is concerned with *approximate, dynamical* noninterference: starting with an initial product state, the interactions that lead to decoherence evolve the state into a superposition of product states each of which evolve approximately without interference,

$$|\Psi(0)\rangle = |S_0\rangle|E_0\rangle \to |\Psi(t)\rangle \approx \sum_i c_i |\tilde{S}_i(t)\rangle|\tilde{E}_j(t)\rangle. \tag{1.7}$$

By inspection we see that states $|\tilde{S}_i(t)\rangle|\tilde{E}_i(t)\rangle$ are eigenstates of the Hamiltonian $\hat{H}$ associated with the Hilbert space. When the number of branches is less then the dimensionality of the system, this implies that the reduced density matrix ρ_S is diagonal in the $\{S_i\}$ basis, and it is consistent to treat ρ_S as a probabilistic ensemble over states $|\tilde{S}_i(t)\rangle$. If the states $|\tilde{S}_i(t)\rangle$ are additionally eigenstates of some classical observables (e.g. position or field value), we say that classicalization has occurred: decoherence has driven a quantum-to-classical transition from a single quantum state to a distribution of classical states. In cosmological applications, these states are generally (approximately) classical spacetime geometries, with a definite metric sourced by some field content. In Everettian terms, the wave function has branched into distinct classical universes. Implicit in this formalism is that decoherence is accompanied by entropy production: from (1.7) we see that the entanglement entropy between system and environment was initially zero but has now increased.

In the context of quantum field theory in curved spacetime, the most common way to define a system is to trace out the degrees of freedom behind some cosmological horizon. Such horizons arise in many contexts: for example in black hole (Schwarzschild, Kerr, Reissner-Nordström) geometries, for accelerated (Rindler) observers in Minkowski space, and in de Sitter space. In all of these geometries, tracing out the degrees of freedom behind the horizon from the ground state of the theory

yields a thermal density matrix, with temperatures

$$T_{\text{Hawking}} = \frac{1}{8\pi GM} = \frac{\kappa}{2\pi}, \qquad T_{\text{Unruh}} = \frac{a}{2\pi}, \qquad T_{\text{Gibbons-Hawking}} = \frac{H}{2\pi}. \qquad (1.8)$$

We can also compute the entanglement entropy across the horizon: for a Rindler observer this is infinite, but in the other two cases we get the famous expression

$$S = \frac{A}{4G} \qquad (1.9)$$

with A the horizon area. Given the accompanying temperature it is natural to identify the entanglement entropy with thermodynamic entropy and treat the horizon as a thermodynamical system in its own right. Furthermore, at least semiclassically both black holes and de Sitter space are states of *maximum* entropy. There are no-hair theorems (for the less-well-known de Sitter case, see [8–11]) which imply in the black hole case that once a horizon forms the state approaches the vacuum solution and in the de Sitter case that perturbations redshift across the horizon. Our universe is manifestly not in a maximum-entropy state, as the existence of this document attests, but in de Sitter phases it approaches one, and if the current observed dark energy is due at least in part to a positive cosmological constant such an equilibrium will be its ultimate fate.

We have therefore identified a number of broad classes of quantum-mechanical constraints on cosmological systems:

- constraints on the Hilbert space decomposition and Hamiltonian to allow for cosmological structure;

- given such structure, constraints on the initial state and Hamiltonian to allow for the emergence of classical spacetimes via decoherence;

- given such classical spacetimes, the need for them to obey gravitational constraints on the entanglement entropy of subsystems.

The last category of constraints fall under the broad category of *holography*; they include the area laws for black holes and de Sitter geometries already mentioned, but also the powerful set of ideas [12–14] relating anti-de Sitter space to conformal field theory in the context of string theory, where the Ryu-Takayanagi formula [15] (which is just (1.9), again!) equates the entropies of regions in the CFT living at the

boundary of AdS with the area of minimal surfaces in the bulk of AdS which share a boundary with these regions.

The papers in this thesis have been organized to begin with explicitly holographic considerations and gradually move towards more general constraints from decoherence on the wave function and Hamiltonian:

- In Chapter 2, we are concerned with the proposed AdS-MERA correspondence, which conjectured that a computationally successful method of approximating the ground states of CFTs on a lattice using tensor networks, the multiscale entanglement renormalization ansatz, should by analogy with AdS/CFT provide a discretization of anti-de Sitter space. Our strategy is to demand that gravitational area laws be obeyed not only for regions ending on the boundary, but also for areas purely inside the bulk–in other words, that the MERA should not allow for regions with entropy larger than that of a black hole of the same size. Imposing such a constraint places severe constraints on the allowed parameters of the MERA and indicates the need to use more general tensor networks to discretize AdS.

- Chapters 3, 4, and 5 comprise a series of three papers which investigate the general-relativistic limit of another proposed holographically-inspired correspondence, this time a much more general one. The ER=EPR correspondence [16] proposes an entropy-area duality for all entangled states: every pair of entangled qubits should be connected by a Planck-scale wormhole. In the general relativistic limit, the conjecture stipulates that two black holes of the same size connected by an Einstein-Rosen bridge (wormhole) should be maximally entangled. Evidence comes from the thermofield double state in AdS/CFT–two noninteracting CFTS on distinct boundaries entangled in a thermal state–which is known to have a bulk description corresponding to maximally entangled Schwarzschild spacetime. If ER=EPR is true, it implies that quantum-mechanical constraints and no-go theorems should map to general-relativistic ones. In Chapter 3 we indeed show that the no-cloning theorem in quantum mechanics is dual to the no-go theorem forbidding topology change in general relativity. In Chapter 4 we verify that just as the presence or absence of entanglement is not observable neither is the presence or absence of a wormhole connected to a given black hole. Finally, in Chapter 5, we show in full generality that we can define an area for a spacetime geometry containing wormholes which is exactly equivalent to the entanglement entropy between

the two sides of the wormhole. In particular we demonstrate that the area is unchanged by operations acting on only a single side of the wormhole, just as entanglement between two systems is unchanged by acting on only one at a time.

- In Chapter 6, we investigate whether the claimed violation of unitarity within black hole complementarity (the "firewall paradox" [261, 262]) can actually be experimentally observed. We propose a slight relaxation of the no-drama condition in black hole complementarity, "little drama," in which macroscopic objects and observables remain unchanged when passing though the black hole horizon but microscopic entanglement structure is diffused throughout the interior just as it is scrambled in the complementary description of the black hole boundary. We show that, after information is diffused, regaining this information in the black hole interior cannot be accomplished before encountering the singularity.

- In Chapter 7, we consider the decoherence properties, and thus the possible branching structures, of late-time de Sitter space. As already anticipated in this Introduction, once the state of the universe approaches pure de Sitter it is exponentially close to a maximum-entropy state and no further entropy production is possible, so further decoherence is forbidden. Since we know the quantum state of the de Sitter vacuum we can determine its branching structure exactly; in particular, it includes no time-dependent fluctuations. If the dimensionality of the Hilbert space of the universe is infinite, as it is in quantum field theory, Poincaré recurrences are also forbidden. Hence under these assumptions the late-time state of a universe with a positive cosmological constant entirely lacks dynamical quantum fluctuations.

- Finally, in Chapter 8 we perform an analysis of the decoherence of inflaton perturbations in slow-roll eternal inflation. Unlike in pure de Sitter space, gravitational-strength interactions suppressed by the slow-roll parameters act to decohere modes in their field value basis once they become sufficiently superhorizon. Starting from the vacuum state, the wave function repeatedly branches as modes progressively grow beyond the threshold of decoherence. Modes smaller than the Hubble scale, or localized observers, therefore experience a cosmological history which varies from branch to branch—that is, the decoherence of large-scale modes backreacts on the subsequent cosmological

evolution. Our analysis of decoherence and backreaction reveals that all of the required conditions necessary for eternal inflation are present in slow-roll inflation.

Consistency Conditions for an AdS/MERA Correspondence

2.1 Introduction

The idea that spacetime might emerge from more fundamental degrees of freedom has long fascinated physicists. The holographic principle suggests that a $(D + 1)$-dimensional spacetime might emerge from degrees of freedom in a D-dimensional theory without gravity [17, 18]. While a completely general implementation of this idea is still lacking, the AdS/CFT correspondence provides a specific example in which to probe the holographic emergence of spacetime. AdS/CFT is a conjectured correspondence between D-dimensional conformal field theories (CFTs) in Minkowski space and $(D + 1)$-dimensional asymptotically anti-de Sitter (AdS) spacetimes [14, 19, 20]. An intriguing aspect of this duality is the Ryu–Takayanagi formula [21, 22], according to which the entanglement entropy of a region B on the boundary is proportional to the area of a codimension-two extremal surface $\tilde{B}$ embedded in the bulk curved spacetime whose boundary is B:

$$S(B) = \frac{\mathrm{area}(\tilde{B})}{4G} + \text{corrections.} \qquad (2.1.1)$$

In other words, given a CFT state, one may think of bulk distance and geometry (at least near the boundary) as being charted out by the entanglement properties of the CFT state.

A central question in this picture of spacetime emerging from entanglement is: What is the precise relationship between bulk degrees of freedom and boundary degrees of freedom? Expressed in a different way, what is the full map between states and operators in the boundary Hilbert space and those in the bulk? While investigations of AdS/CFT have thrown a great deal of light on this question, explicit simple models are still very helpful for studying it in more detail.

Meanwhile, from a very different perspective, tensor networks have arisen as a useful way to calculate quantum states in strongly-interacting many-body systems [23]. One significant example is the Multi-scale Entanglement Renormalization Ansatz (MERA) [24], which is relevant for critical (gapless) systems, *i.e.*, CFTs. Starting from a simple state in a low-dimensional Hilbert space, acting repeatedly with fixed tensors living on a network lattice produces an entangled wave function for the quantum system of interest; varying with respect to the tensor parameters efficiently computes the system's ground state.

Working "backwards" in the MERA, starting with the ground state and gradually removing entanglement, produces a set of consecutively renormalized quantum states.

This process reveals a renormalization direction along the graph, which may be thought of as an emergent radial direction of space. As pointed out by Swingle [25], the MERA graph can serve as a lattice discretization of spatial slices of AdS. Furthermore, one can use the MERA to calculate the entanglement entropy of regions of the original (boundary) critical system; this calculation amounts to tracing over bonds in the tensor network that cross the causal cone of the boundary region. The causal cone is a sort of extremal surface for the MERA, motivating comparison to the Ryu–Takayanagi formula.

It is therefore natural to conjecture that the MERA provides a concrete implementation of the emergence of spacetime, in the form of a correspondence between boundary and bulk regions reminiscent of AdS/CFT [25]. Such an AdS/MERA correspondence would be extremely useful, since the basic building blocks of the MERA are discrete quantum degrees of freedom from which quantities of physical interest may be directly calculated. Some specific ideas along these lines have recently been investigated [26–29].

In this paper, we take a step back and investigate what it would mean for such a correspondence to exist and the constraints it must satisfy in order to recover properties we expect of physics in a bulk emergent spacetime. After reviewing the MERA itself and possible construals of the AdS/MERA correspondence in the next section, in Sec. 2.3 we then derive relationships between the MERA lattice and the geometry of AdS. We find that the MERA is unable to describe physics on scales shorter than the AdS radius. In Sec. 5.2 we explore constraints from calculating the entanglement entropy of regions on the boundary, in which we are able to relate MERA parameters to the central charge of the CFT. Finally, in Sec. 3.3 we apply the covariant entropy (Bousso) bound to regions of the bulk lattice. In the most naïve version of the AdS/MERA correspondence, we find that no combination of parameters is consistent with this bound, but we suggest that generalizations of the tensor network may be able to provide a useful correspondence.

2.2 AdS/MERA

Let us begin by recalling the definition and construction of the MERA. We will then introduce the AdS/MERA correspondence and discuss the motivation for and consequences of this proposal.
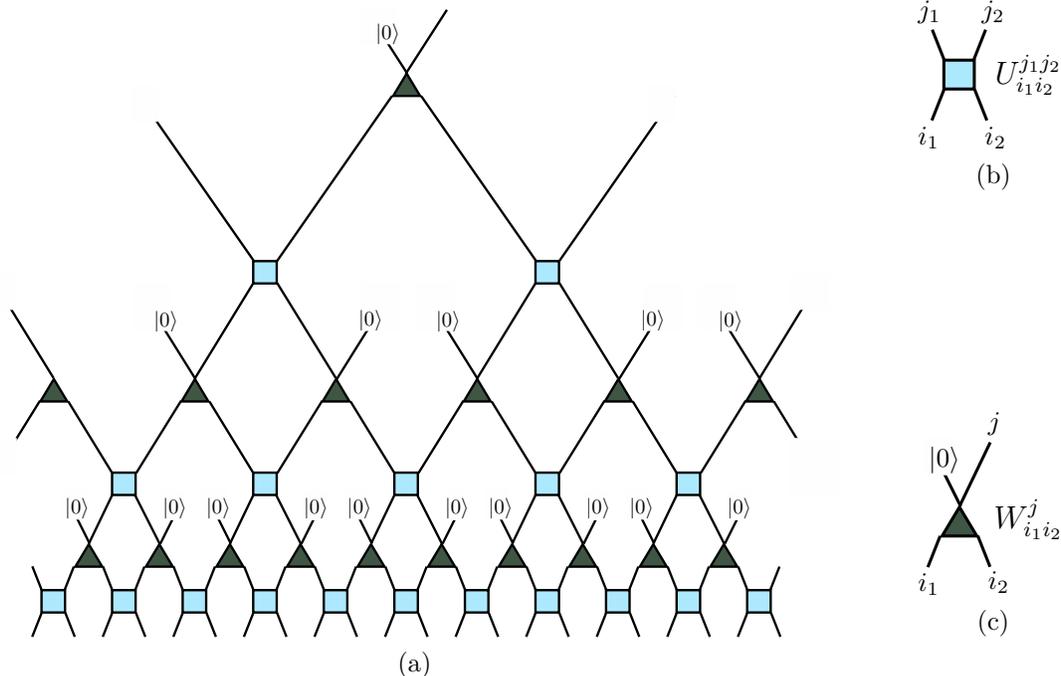
Figure 2.1. (a) Basic construction of a $k = 2$ MERA (2 sites renormalized to 1). (b) The squares represent disentanglers: unitary maps that, from the moving-upward perspective, remove entanglement between two adjacent sites. (c) The triangles represent isometries: linear maps that, again from the moving-upward perspective, coarse-grain two sites into one. Moving downward, we may think of isometries as unitary operators that, in the MERA, map a state in $V \otimes |0\rangle$ into $V \otimes V$. The i and j labels in (b) and (c) represent the tensor indices of the disentangler and isometry.

2.2.1 Review of the MERA

The MERA is a particular type of tensor network that provides a computationally efficient way of finding the ground states of critical quantum many-body systems, *i.e.* CFTs, in D dimensions. (For a recent review of tensor networks in general, see Ref. [23]. Detailed analyses of the MERA are given in [24, 30, 31] and references therein.) In this work, we restrict our attention to the case $D = 1 + 1$.

The MERA tensor network is shown in Fig. 2.1. The quantum system being modeled by the MERA lives at the bottom of the diagram, henceforth "the boundary" in anticipation of the AdS/MERA connection to be explored later. We can think of the tensor network as a quantum circuit that either runs from the top down, starting

with a simple input state and constructing the boundary state, or from the bottom up, renormalizing a boundary state via coarse-graining. One defining parameter of the MERA is the rescaling factor k, defining the number of sites in a block to be coarse-grained; in Fig. 2.1 we have portrayed the case $k = 2$. The squares and triangles are the tensors: multilinear maps between direct products of vector spaces. Each line represents an index i of the corresponding tensor, ranging over values from 1 to the "bond dimension" χ. The boundary Hilbert space $\mathcal{H}_{\text{boundary}} = V^{\otimes \mathcal{N}_{\text{boundary}}}$ is given by a tensor product of $\mathcal{N}_{\text{boundary}}$ individual spaces V, each of dimension χ. (In principle the dimension of the factors in the boundary could be different from the bond dimension of the MERA, and indeed the bond dimensions could vary over the different tensors. We will assume these are all equal.)

As its name promises, the MERA serves to renormalize the initial boundary state via coarse-graining. If we were to implement the MERA for only a few levels, we would end up with a quantum state in a smaller Hilbert space (defined on a fixed level of the tensor network), retaining some features of the original state but with some of the entanglement removed. However, we can also run the MERA backwards, to obtain a boundary state from a simple initial input. By varying the parameters in the individual tensors, we can look for an approximation of the ground state of the CFT on the boundary. Numerical evidence indicates that this process provides a computationally efficient method of constructing such ground states [31, 32].

The tensors, or gates, of the MERA come in two types. The first type are the disentanglers, represented by squares in Fig. 2.1. These are unitary maps $U : V \otimes V \to V \otimes V$, as in Fig. 2.1b. The name comes from thinking of moving upward through the network, in the direction of coarse-graining, where the disentanglers serve to remove local entanglement; as we move downward, of course, they take product states and entangle them. The second type of tensors are the isometries, represented by triangles. From the moving-downward perspective these are linear maps $W : V \to V \otimes V$; moving upward, they implement the coarse-graining, see Fig. 2.1c. The isometries are subject to the further requirement that $W^\dagger W = I_V$, where I_V is the identity map on V, and $WW^\dagger = P_A$, where P_A is a projector onto some subspace $A \subset V \otimes V$. From the top-down perspective, we can also think of the isometries as bijective unitary operators $W_U : V \otimes V \to V \otimes V$, for which a fixed "ancilla" state (typically the ground state $|0\rangle$) is inserted in one of the input factors, as shown in Fig. 2.1c. More generally, isometries could map $q < k$ sites onto k sites, $W : V^{\otimes q} \to V^{\otimes k}$.

The MERA is not the simplest tensor network which implements coarse-graining.

For instance, the tree tensor network [33] (also considered in a holographic context in Ref. [26]), similar to MERA but without any disentanglers, also implements coarse-graining. However, tensor networks without disentanglers fail to capture the physics of systems without exponentially-decaying correlations, and consequently cannot reproduce a CFT ground state.

An example that invites analysis with a MERA is the transverse-field Ising model [34]. In $1+1$ dimensions, the model describes a chain of spins with nearest-neighbor interactions subject to a transverse magnetic field. Its Hamiltonian is

$$\hat{H} = -J \sum_i \hat{\sigma}_i^z \hat{\sigma}_{i+1}^z - h \sum_i \hat{\sigma}_i^x \,, \tag{2.2.1}$$

where $\hat{\sigma}_i^z$ and $\hat{\sigma}_i^x$ are Pauli operators and where J and h set the strength of the nearest-neighbor interactions and the magnetic field, respectively. Notably, the system achieves criticality at $J = h$, where a quantum phase transition occurs between ordered ($J > h$) and disordered ($J < h$) phases. In this example, the open legs at the bottom of the MERA describe the state of the one-dimensional lattice of spins. A single application of disentanglers and isometries can be thought of as a true real-space renormalization, producing a lattice of spins that is less dense than the preceding lattice by a factor of q/k.

In general, much information is required to describe an arbitrary MERA. In principle, the Hilbert spaces, the disentanglers, and the isometries could all be different. Also, for $k > 2$, there is no canonical way of laying out the disentanglers and isometries; the circuit itself must be specified. We will restrict ourselves to the case $q = 1$, so that isometries have 1 upward-going leg and k downward-going legs. Further, without loss of generality, we take the same vector spaces, disentanglers, and isometries everywhere in the MERA, a simplification that is enforced by the symmetries of the boundary ground state. These symmetries — namely, translation- and scale-invariance — dictate that the MERA parameters and structure be homogeneous across the whole tensor network.

For geometric considerations, it is useful to abstract away all of the information about unitary operators and to draw a MERA as a graph as shown in Fig. 2.2. In such a graph, we only indicate the connectivity of sites at any given level of coarse-graining as well as the connectivity of sites under renormalization group flow.
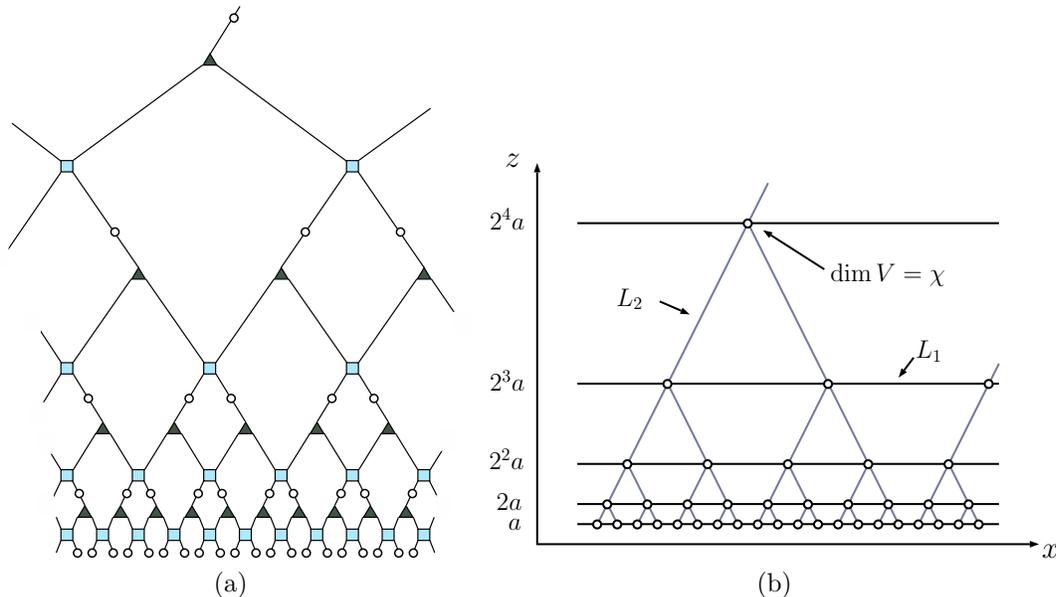
Figure 2.2. (a) A $k = 2$ MERA, and (b) the same MERA with its disentanglers and isometries suppressed. The horizontal lines in the graph on the right indicate lattice connectivity at different renormalization depths, and the vertical lines indicate which sites at different depths are related via coarse-graining due to the isometries. Each site, represented by a circle, is associated with a Hilbert space V with bond dimension χ. In the simplest case, a copy of the same Hilbert space is located at each site. When assigning a metric to the graph on the right, translation and scale invariance dictate that there are only two possible length scales: a horizontal proper length L_1 and a vertical proper length L_2.

2.2.2 An AdS/MERA correspondence?

The possibility of a correspondence between AdS and the MERA was first proposed by Swingle in Ref. [25], where it was noted that the MERA seems to capture certain key geometric features of AdS. At the most basic level, when viewed as a graph with legs of fixed length, a MERA may be thought of as a discretization of the hyperbolic plane, which is a spatial slice of AdS$_3$. In this discretization, the base of the MERA tree lies on the boundary of the AdS slice and the MERA lattice sites fill out the bulk of the slice [25, 35].

Interestingly, the structure of a MERA is such that it seems to go beyond a simple discretization of the hyperbolic plane. Certain discrete paths in the MERA naturally reproduce geodesics of the hyperbolic plane [25, 36]. Moreover, this phenomenon makes it possible to understand the computation of CFT entanglement entropy using a MERA as a discrete realization of the Ryu–Takayanagi formula [37]. These and

other examples [25, 36] seem to suggest that a MERA may in fact be elucidating the structural relationship between physics on the boundary of AdS and its bulk.

In this work we take the term "AdS/MERA correspondence" to mean more than simply a matching of graph geometry and continuous geometry. In the spirit of the AdS/CFT correspondence, we suppose that (at least some aspects of) both boundary and bulk physics are described by appropriate Hilbert spaces $\mathcal{H}_{\text{boundary}}$ and $\mathcal{H}_{\text{bulk}}$ respectively, which must have equal dimensions. A full AdS/MERA correspondence would then be a specification of these Hilbert spaces, as well as a prescription which makes use of the MERA to holographically map states and operators in $\mathcal{H}_{\text{boundary}}$ to corresponding states and operators in $\mathcal{H}_{\text{bulk}}$ and vice-versa. To preserve locality in the bulk and the symmetries of AdS, it is natural to identify $\mathcal{H}_{\text{bulk}}$ with the tensor product of individual spaces V_{bulk}, each located at one site of the MERA. If it exists, this correspondence provides a formulation of bulk calculations in terms of the MERA. An AdS/MERA correspondence should allow us to, for example, calculate bulk correlation functions, or bulk entanglement entropies using tools from or the structure of the MERA.

There is one straightforward way to construct such a map $\mathcal{H}_{\text{boundary}} \leftrightarrow \mathcal{H}_{\text{bulk}}$. We have noted that the isometries $W : V \rightarrow V \otimes V$ can be thought of as unitaries $W_U : V \otimes V \rightarrow V \otimes V$ by imagining that a fixed ancillary state $|0\rangle$ is inserted in the first factor; for a k-to-one MERA, one would insert $k - 1$ copies of the $|0\rangle$ ancilla at each site to unitarize the isometries. From that perspective, running upwards in the tensor network provides a map from the MERA ground state on the boundary to a state $|0\rangle^{\otimes(k-1)\mathcal{N}_{\text{bulk}}} \in V^{\otimes(k-1)\mathcal{N}_{\text{bulk}}}$, where at each isometry there is a copy of $V^{\otimes(k-1)}$ and $\mathcal{N}_{\text{bulk}}$ denotes the number of bulk lattice sites, excluding the boundary layer. As we ultimately show in Sec. 3.3, one has $\mathcal{N}_{\text{boundary}} = (k - 1)\mathcal{N}_{\text{bulk}}$. We can then identify $\mathcal{H}_{\text{boundary}} = \mathcal{H}_{\text{bulk}} = V^{\otimes\mathcal{N}_{\text{boundary}}}$ and think of the tensor network as a quantum circuit providing a map between arbitrary states $\mathcal{H}_{\text{boundary}} \rightarrow \mathcal{H}_{\text{bulk}}$. In this construction, the MERA ground state on the boundary gets mapped to the factorized bulk state $|0\rangle^{\otimes(k-1)\mathcal{N}_{\text{bulk}}}$, but other boundary states will in general produce entangled states in the bulk (keeping the tensors themselves fixed).

Something very much like this construction was proposed by Qi [26], under the name "Exact Holographic Mapping" (EHM). That work examined a tensor network that was not quite a MERA, as no disentanglers were included, only isometries. As a result, while there is a map $\mathcal{H}_{\text{boundary}} \rightarrow \mathcal{H}_{\text{bulk}}$, the boundary state constructed by the tensor network does not have the entanglement structure of a CFT ground state. In particular, it does not seem to reproduce the Ryu–Takayanagi formula in a

robust way. Alternatively, we can depart from Qi by keeping a true MERA with the disentanglers left in, in which case the bulk state constructed by the quantum circuit has no entanglement: it is a completely factorized product of the ancilla states. Such a state doesn't precisely match our expectation for what a bulk ground state should look like, since there should be at least some entanglement between nearby regions of space.

Therefore, while it is relatively simple to imagine constructing a bulk Hilbert space and a map between it and the boundary Hilbert space, it is not straightforward to construct such a map that has all of the properties we desire. It might very well be possible to find such a construction, either by starting with a slightly different boundary state, or by adding some additional structure to the MERA.

For the purposes of this paper we will be noncommittal. That is, we will imagine that there is a bulk Hilbert space constructed as the tensor product of smaller spaces at each MERA site, and that there exists a map $\mathcal{H}_{\mathrm{boundary}} \rightarrow \mathcal{H}_{\mathrm{bulk}}$ that can be constructed from the MERA, but we will not specify precisely what that map might be. We will see that we are able to derive bounds simply from the requirements that the hypothetical correspondence should allow us to recover the properties we expect of bulk physics, including the background AdS geometry and features of semiclassical quantum gravity such as the Bousso bound on bulk entropy.

2.3 MERA and Geometry

If a MERA is a truly geometrical object that describes a slice of AdS, then the graph geometry of a MERA should give the same answers to geometric questions as the continuous geometry of a slice of AdS. Here, we reconsider the observation by Swingle [25, 36] that certain trajectories on the MERA coincide with trajectories in AdS and we investigate the constraints that this correspondence places on the graph metric of the MERA. We find that a MERA necessarily describes geometry on super-AdS length scales, moreover, there is no redefinition of the MERA coordinates that results in the proper distance between MERA sites mapping to any sub-AdS length scale.

2.3.1 Consistency conditions from matching trajectories

In order to speak of graph geometry, one must put a metric on the MERA graph, *i.e.*, one must assign a proper length to each bond in the graph of Fig. 2.2. Presumably, the metric should originate from correlations between the sites in the MERA. In the absence of an explicit identification of the origin of the graph metric, however, at

least in the case of a MERA describing the ground state of a CFT, it is sensible to identify two length scales. Explicitly, we must assign a proper length L_1 to horizontal bonds and a proper length L_2 to vertical bonds. Indeed, translational and conformal invariance guarantee that these are the only two length scales in any graph metric one can assign to a MERA for which an AdS/MERA correspondence exists. In particular, the ground state of a CFT is translation invariant, so each horizontal bond in the finest (UV-most) lattice should have the same proper length so as to respect this symmetry. Self-similarity at all scales then requires that any horizontal bond at any level of renormalization have this same proper length. There is no *a priori* reason why the vertical bonds should share the proper length of the horizontal bonds and indeed we will see that their proper length will be different. However, again by self-similarity and translation invariance, all vertical bonds must be assigned the same proper length.

The observation in Ref. [25] that certain paths in the MERA graph coincide with corresponding paths in slices of AdS is what established the possibility of an AdS/MERA correspondence. Here we will carefully examine these paths and determine what constraints the requirements that they match place on MERA parameters, *i.e.*, on the bond lengths L_1 and L_2 and on the rescaling factor k.

Consider a constant-time slice of AdS$_3$ with the following metric:

$$\mathrm{d}s^2 = \frac{L^2}{z^2}(\mathrm{d}z^2 + \mathrm{d}x^2). \tag{2.3.1}$$

We will compare the proper lengths of straight horizontal lines and geodesics in the AdS slice to the proper lengths of the corresponding paths in the MERA graph. In the AdS slice, let γ_1 be a straight horizontal line ($\mathrm{d}z = 0$) sitting at $z = z_0$ with coordinate length x_0. Let γ_2 be a geodesic whose endpoints lie near the boundary $z = 0$ and are separated by a coordinate distance x_0 at the boundary. In this choice of coordinates, such a geodesic looks like a semicircle (see Fig. 2.3). It is a straightforward computation to show that the proper lengths of these curves are

$$|\gamma_1|_{\mathrm{AdS}} = \frac{L}{z_0}x_0 \qquad \text{and} \qquad |\gamma_2|_{\mathrm{AdS}} = 2L \ln\left(\frac{x_0}{a}\right). \tag{2.3.2}$$

Note that there is a UV cutoff at $z = a \ll x_0$ and that we have neglected terms of order a/x_0.

We fix L_1 and L_2 by comparing γ_1 and γ_2 to horizontal lines and "geodesics" in the MERA, respectively. Consider two sites in a horizontal lattice at depth m (*i.e.*,
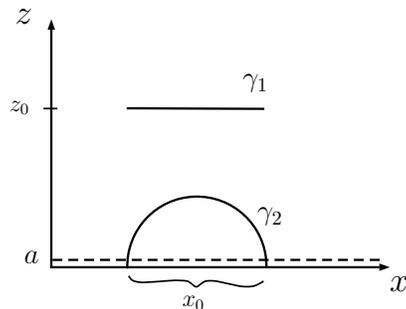
Figure 2.3. A horizontal line (γ_1) and a geodesic (γ_2) in a spatial slice of AdS$_3$.

m renormalizations of the UV-most lattice) and separated by a coordinate distance x_0 in the coordinate system shown in Fig. 2.2. By fiat, this lattice sits at $z_0 = k^m a$. The number of bonds between the two sites at depth m is $x_0/(k^m a)$ (see Fig. 2.2 for the case $k = 2$). It follows that the proper length of the line connecting the two points is just

$$|\gamma_1|_{\text{MERA}} = L_1 \cdot (\text{number of bonds between endpoints})$$
$$= L_1 \left.\frac{x_0}{z_0}\right|_{z_0 = k^m a}.$$

(2.3.3)

To have $|\gamma_1|_{\text{AdS}} = |\gamma_1|_{\text{MERA}}$, we should therefore set $L_1 = L$.

Similarly, consider two lattice sites on the UV-most lattice separated by a coordinate distance x_0. If we assume that $x_0 \gg a$, then the shortest path (geodesic) in the MERA connecting the two lattice sites is the path that goes up in the renormalization direction and then back down again. The two sites are separated by x_0/a bonds on the UV-most lattice, so $\log_k(x_0/a)$ renormalization steps are needed to make the sites either adjacent or superimposed. This means that the geodesic that connects the endpoints is made up of $2\log_k(x_0/a)$ bonds (as we have to go up and then back down again, giving the factor of 2). It follows that the proper length of the geodesic is

$$|\gamma_2|_{\text{MERA}} = L_2 \cdot (\text{number of bonds in the geodesic})$$
$$= 2L_2 \log_k\left(\frac{x_0}{a}\right).$$

(2.3.4)

To have $|\gamma_2|_{\text{AdS}} = |\gamma_2|_{\text{MERA}}$, we should therefore set $L_2 = L \ln k$.

2.3.2 Limits on sub-AdS scale physics

One aspect of the matching of geodesics that is immediately apparent is that the MERA scales L_1 and L_2 that parametrize the proper distance between lattice sites are of order the AdS scale L or larger, as was also noted in Refs. [25, 35]. This runs counter to the typical expectation that, in a discretization of spacetime, one expects the granularity to be apparent on the UV, rather than the IR, scale. That is, sub-AdS scale locality is not manifested in the MERA construction and must be encoded within each tensor factor [36].

One could try to evade this difficulty by attempting to redefine the MERA coordinates $(x, z)^{\mathrm{MERA}}$ (those of Fig. 2.2) as functions of the AdS coordinates $(x, z)^{\mathrm{AdS}}$ (those of Fig. 2.3) and taking a continuum limit; above, we assumed that the two sets of coordinates were simply identified. That is, suppose $x^{\mathrm{MERA}} = f(x^{\mathrm{AdS}})$ and $z^{\mathrm{MERA}} = g(z^{\mathrm{AdS}})$. (For example, one could consider $f(x) = \varepsilon x$ for small ε and imagine taking the continuum limit, with the aim of making L_1 much smaller than the AdS scale.) If a is still the UV cutoff on the AdS side, then in the MERA we have $f(a)$ as the UV-most lattice spacing and $g(a)$ as the UV cutoff in the holographic direction. Consider the computation of $|\gamma_1|$. From the AdS side, we have $|\gamma_1|_{\mathrm{AdS}} = L x_0^{\mathrm{AdS}}/z_0^{\mathrm{AdS}}$. On the MERA side, the number of sites spanned by $x_0^{\mathrm{MERA}} = f(x_0^{\mathrm{AdS}})$ is $x_0^{\mathrm{MERA}}/k^m f(a)$, while the holographic coordinate is $z_0^{\mathrm{MERA}} = k^m g(a)$. Hence,

$$|\gamma_1|_{\mathrm{MERA}} = L_1 \frac{f(x_0^{\mathrm{AdS}})}{f(a)} \frac{g(a)}{g(z_0^{\mathrm{AdS}})}. \tag{2.3.5}$$

Equating $|\gamma_1|_{\mathrm{AdS}} = |\gamma_1|_{\mathrm{MERA}} \equiv |\gamma_1|$, we have

$$g(z_0^{\mathrm{AdS}}) \frac{\partial}{\partial x_0^{\mathrm{AdS}}} |\gamma_1| = L_1 \frac{f'(x_0^{\mathrm{AdS}})}{f(a)} g(a) = L \frac{g(z_0^{\mathrm{AdS}})}{z_0^{\mathrm{AdS}}}. \tag{2.3.6}$$

Since the right side of the first equality only depends on x_0^{AdS} and the second equality only depends on z_0^{AdS}, but we can vary both parameters independently, both expressions must be independent of both AdS coordinates. Hence, we must have $f(x) = \varepsilon_x x$ and $g(z) = \varepsilon_z z$ for some constants ε_x and ε_z. Plugging everything back into Eq. (2.3.5) and comparing with $|\gamma_1|_{\mathrm{AdS}}$, we again find that $L_1 = L$, so no continuum limit is possible. Similarly, in computing $|\gamma_2|$, we note that the number of bonds between the endpoints on the UV-most lattice level is $x_0^{\mathrm{MERA}}/f(a)$, so the geodesic connecting the endpoints has $2\log_k(x_0^{\mathrm{MERA}}/\varepsilon_x a)$ bonds. On the other hand,

we have $|\gamma_2|_{\text{AdS}} = 2L \ln(x_0^{\text{AdS}}/a) = 2L \ln(x_0^{\text{MERA}}/\varepsilon_x a)$. That is, in equating $|\gamma_2|_{\text{AdS}}$ and $|\gamma_2|_{\text{MERA}}$, we must again set $L_2 = L \ln k$. We thus also find that no continuum limit is possible in the holographic direction. That is, we have shown that there is a constant normalization freedom in the definition of each of the coordinate distances on the AdS and MERA sides of any AdS/MERA duality, but such a coordinate ambiguity is unphysical and does not allow one to take a continuum limit. One still finds that the physical MERA parameters L_1 and L_2 are AdS scale. This means that there truly is no sense in which a discrete MERA can directly describe sub-AdS scale physics without the addition of supplemental structure to replace the individual tensors. This fact limits the ability of the MERA to be a complete description of the gravity theory without such additional structure. It might be the case that one needs a field theoretic generalization of the MERA, such as continuous MERA (cMERA) [38–40] or some local expansion of the individual tensors into discrete tensor networks with a different graph structure to describe sub-AdS physics, but such a significant generalization of the tensor network is beyond the scope of this work and in any case would no longer correspond to a MERA proper.

2.4 Constraints from Boundary Entanglement Entropy

Because the MERA can efficiently describe critical systems on a lattice, quantities computed in the MERA on scales much larger than the lattice spacing should agree with CFT results. In this section, we will compute the entanglement entropy of ℓ_0 contiguous sites in the MERA and exploit known CFT results to obtain constraints on the properties of the MERA. In particular, we will find an inequality relating the MERA rescaling factor k and bond dimension χ to the CFT central charge c. This constraint is interesting in its own right, but it will prove critical in the next section when we begin to compute bulk properties.

2.4.1 MERA and CFT entanglement entropy

For a $(1+1)$-dimensional CFT in a pure state, the von Neumann entropy of a finite interval B, which is typically referred to as the entanglement entropy, is known to be [41, 42]

$$S(B) = \frac{c}{3} \ln \ell_0 \,, \tag{2.4.1}$$

where the length of the interval is much smaller than the system size. Here, ℓ_0 is the length of the interval in units of the UV cutoff. In the notation of the last section, we

have $\ell_0 = x_0/a$. In the special case that the CFT is dual to AdS in $2+1$ dimensions, the central charge is set by the Brown–Henneaux formula [43],

$$c = \frac{3L}{2G}. \tag{2.4.2}$$

Also note that the length of the geodesic that connects the two ends of B (the curve γ_2 in Fig. 2.3) is given in Eq. (2.3.2) by $|\gamma_2| = 2L \ln \ell_0$. The Brown–Henneaux relation allows us to reproduce the Ryu–Takayanagi formula [21, 44] from the entanglement entropy,

$$S(B) = \frac{\text{area}(\tilde{B})}{4G}, \tag{2.4.3}$$

where $\tilde{B} = \gamma_2$ is the extremal bulk surface with the same boundary as B. For a boundary with one spatial dimension and a bulk with two spatial dimensions, any simply-connected region B is an interval, the extremal bulk surface is a geodesic, $\text{area}(\tilde{B})$ is a length, and G has mass dimension -1.

The MERA calculation of the entanglement entropy of ℓ_0 sites in the CFT has an analogous geometric interpretation. Suppose one is given the MERA representation of a lattice CFT ground state, *i.e.*, one uses a MERA to generate the CFT state. Denote by $S_{\text{MERA}}(\ell_0)$ the entanglement entropy of the resulting state restricted to ℓ_0 sites. In Ref. [37], it is shown that for a specific, optimal choice of ℓ_0 sites, for ℓ_0 parametrically large, the following bound is placed on $S_{\text{MERA}}(\ell_0)$ for a MERA with $k = 2$:

$$S_{\text{MERA}}(\ell_0) \leq 2 \log_2 \ell_0 \ln \chi. \tag{2.4.4}$$

Parsing the equation above, this bound essentially counts the number of bonds that the *causal cone* of the ℓ_0 sites in question crosses ($\sim 2 \log_2 \ell_0$) and $\ln \chi$ is the maximum entanglement entropy that a single bond can possess when the rest of the MERA is traced out.

The causal cone of a region B consisting of ℓ_0 contiguous UV sites in a MERA resembles a bulk extremal surface for the boundary region B. Given ℓ_0 sites in the UV, their causal cone is defined as the part of the MERA on which the reduced density matrix (or in other words, the state) of B depends. An example of a causal cone is illustrated in Fig. 2.4.

In particular, note that the number of bonds that a causal cone crosses up to any fixed layer scales like the length of the boundary of the causal cone up to that layer. It is in this sense that Eq. (2.4.4) is a MERA version of Ryu–Takayanagi.
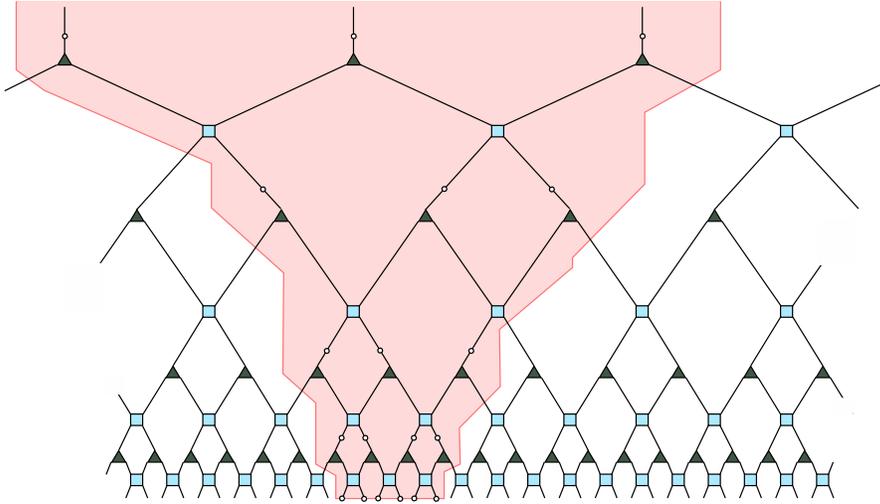
Figure 2.4. Causal cone (shaded) for a set of $\ell_0 = 6$ sites in a MERA with $k = 2$. The width ℓ_m of the causal cone at depth m is $\ell_1 = 4$, $\ell_2 = 3$, $\ell_3 = 3$, $\ell_4 = 3$, etc. The crossover scale for this causal cone occurs at $\bar{m} = 2$. Between the zeroth and first layer, $n_1^{\mathrm{tr}} = 2$ bonds are cut by the causal cone. Similarly, $n_2^{\mathrm{tr}} = 2$, $n_3^{\mathrm{tr}} = 3$, etc.

Also note that the width of the causal cone shrinks by a factor of $\sim 1/k$ after every renormalization step until its width is comparable to k. As such, if one denotes the width of the causal cone at a layer m by ℓ_m, then ℓ_m is roughly constant for all m greater than some $\bar{m}$ (see Fig. 2.4). The scale $\bar{m}$ is called the *crossover scale*.

For general k, it is also possible to formulate a bound similar to Eq. (2.4.4) for the entanglement entropy of ℓ_0 sites. For parametrically large ℓ_0, we find that

$$S_{\mathrm{MERA}}(\ell_0; B) \leq 4(k-1)\log_k \ell_0 \ln \chi. \tag{2.4.5}$$

We demonstrate this bound in App. 2.A using techniques that are similar to those developed in Ref. [37]. In particular, note that we do not allow ourselves to choose the location of the ℓ_0 sites in question. As such, we remind ourselves that S_{MERA} can depend on the location of the region B (and not only its size) by including it in the argument of S_{MERA}. This is also the reason why our Eq. (2.4.5) is more conservative than the optimal bound given in Eq. (2.4.4).

2.4.2 Constraining S_{MERA}

Let us examine Eq. (2.4.5) a bit more closely. As discussed in App. 2.A, $4(k-1)$ is an upper bound on the number of bonds that the causal cone could cut at any given depth m below the crossover scale $\bar{m}$. (The crossover scale $\bar{m}$ is attained after roughly $\log_k \ell_0$ renormalization steps.) For a given causal cone, *i.e.*, for ℓ_0 sites at a given location with respect to the MERA, let us parametrize our ignorance by writing

$$S_{\text{MERA}}(\ell_0; B) \leq 4f_B(k) \log_k \ell_0 \ln \chi \,, \qquad (2.4.6)$$

where $f_B(k)$ grows no faster than $(k-1)$ and counts the (average) number of bonds cut by the causal cone at any depth up to the crossover scale. Explicitly,

$$f_B(k) \equiv \frac{1}{4\bar{m}} \sum_{m=0}^{\bar{m}-1} n_m^{\text{tr}} \,, \qquad (2.4.7)$$

where n_m^{tr} denotes the number of bonds that the causal cone cuts at the mth level.

Each cut bond contributes at most $\ln \chi$ to the entropy (the case of maximal entanglement). As such, it is instructive to introduce a parameter $\eta_B \in [0, 1]$ that describes the degree of entanglement of the sites in the causal cone. In doing so we may rewrite the inequality (2.4.6) as an equality:

$$S_{\text{MERA}}(\ell_0; B) = 4f_B(k) \log_k \ell_0 \cdot \eta_B \ln \chi. \qquad (2.4.8)$$

The quantity $\eta_B \ln \chi$ is the average entanglement entropy per cut bond in the causal cone of B. Equivalently, Eq. (2.4.8) may be taken as the definition of η_B.

This definition of η_B of course depends on the location of B and only applies to bonds that are cut by the causal cone of B. In what follows, it will be advantageous to have a notion of average entanglement entropy per bond that applies to *all* bonds in the MERA. To this end, start with a lattice consisting of ℓ_{tot} sites in total and consider the limit in which the size of a region B is unbounded but where the ratio ℓ_0/ℓ_{tot} is held constant (so that B does not grow to encompass the whole domain of the CFT). In this limit, $S_{\text{MERA}}(\ell_0; B) \to S_{\text{MERA}}(\ell_0)$ and $f_B(k) \to f(k)$ should be independent of the exact location of B, *i.e.*, S_{MERA} should exactly agree with Eq. (2.4.1). Let us consequently define the average entanglement entropy per bond in the MERA:

$$\eta \ln \chi = \lim_{\ell_0 \to \infty} \frac{S_{\text{MERA}}(\ell_0)}{4f(k) \log_k(\ell_0)} \,, \qquad (2.4.9)$$

The quantity η is then a property of the MERA itself.

Intuitively, one would not expect each individual bond in the MERA to be maximally entangled and so it should be possible to constrain η more tightly than $\eta \leq 1$. This expectation is made more precise via the following considerations. To begin, consider a MERA with $k = 2$ and examine a pair of isometries at a fixed depth m. As indicated in Fig. 2.5a, let ρ_2 denote the density matrix of the bonds and ancillae emanating from the two isometries and let ρ_1 denote the density matrix of the four highlighted bonds below the isometries. We again assume that the ancillae are initialized to the pure product state composed of factors of $|0\rangle$. Taking into account the ancillae, or in other words promoting the isometries to unitaries, we see that ρ_1 and ρ_2 are related by a unitary transformation, so $S(\rho_1) = S(\rho_2)$. By assumption, the state of each ancilla is $|0\rangle$, so $\rho_2 = \tilde{\rho}_2 \otimes |0\rangle\langle 0| \otimes |0\rangle\langle 0|$ for some density matrix $\tilde{\rho}_2$. This in turn implies that $S(\rho_2) = S(\tilde{\rho}_2) \leq 2 \ln \chi$. From the definition of η above, the entanglement entropy of a single bond is asymptotically given by $\eta \ln \chi$, so $S(\rho_1) \simeq 4\eta \ln \chi$. It therefore follows that $\eta \leq 1/2$.
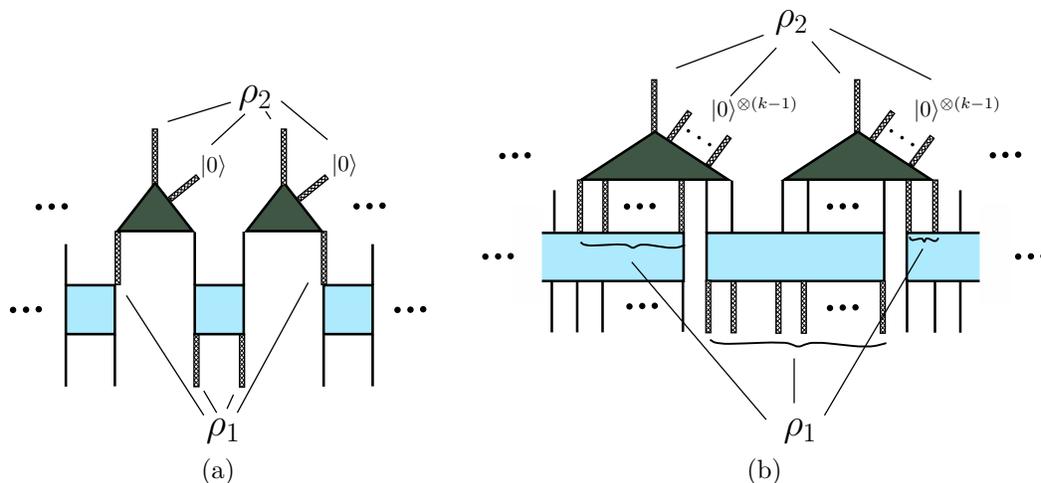
Figure 2.5. A pair of isometries with their ancillae explicitly indicated for a MERA with (a) $k = 2$ and (b) general k. The thick bonds below the isometries, the state of which is denoted by ρ_1, are unitarily related to the bonds that exit the isometries and the ancillae, the state of which is denoted by ρ_2.

For general k, the argument is nearly identical. We again begin by considering a pair of isometries at a given level m (see Fig. 2.5b). Analogously with the $k = 2$ case, let ρ_2 denote the density matrix of the two bonds and $2k - 2$ ancillae emanating from the two isometries and let ρ_1 denote the density matrix of the $2k$ highlighted

bonds below the isometries. There is only one disentangler that straddles both of the isometries in question for any layout of the MERA. As such, at most k of the lower bonds enter a disentangler from below and the rest directly enter the isometries. Here as well ρ_1 and ρ_2 are related by a unitary transformation so that $S(\rho_1) = S(\rho_2)$. Similarly, $\rho_2 = \tilde{\rho}_2 \otimes (|0\rangle\langle 0|)^{\otimes 2k-2}$ for some density matrix $\tilde{\rho}_2$, so $S(\rho_2) = S(\tilde{\rho}_2) \leq 2 \ln \chi$. The region described by ρ_1 always consists of $2k$ bonds, so we may again asymptotically write $S(\rho_1) \simeq 2k\eta \ln \chi$. It therefore follows that $k\eta \leq 1$, and since $f(k) \leq (k-1)$, we may write

$$\eta f(k) \leq \frac{k-1}{k} \,. \tag{2.4.10}$$

We note that, in computational practice, one typically does not use the "worst-case scenario" construction explored in App. 2.A; a more conventional construction would result in a tighter bound on $f(k)$ and hence a stricter inequality than Eq. (2.4.10). For our purposes, however, we will remain as conservative as possible and therefore use the inequality (2.4.10) in our subsequent bounds.

2.4.3 Matching to the CFT

Finally, we obtain a constraint on k, χ, and η in terms of the central charge c by collecting the results of this section. Let us work in the limit where the interval is much larger than the lattice spacing, $\log_k \ell_0 \gg 1$. We have seen that this is precisely the regime in which η and $f(k)$ are well-defined quantities independent of the choice of B. It is also the regime in which we can equate the CFT entropy $S(\ell_0) = (c/3) \ln \ell_0$ with the MERA entropy (2.4.8). Doing so, the central charge is given by

$$c = \frac{3L}{2G} = 12\eta \, f(k) \frac{\ln \chi}{\ln k} \,. \tag{2.4.11}$$

Then in light of Eq. (2.4.10), we find that

$$c \leq 12 \left(\frac{k-1}{k \ln k}\right) \ln \chi \,. \tag{2.4.12}$$

To recapitulate, given a CFT with central charge c and a MERA representation of its ground state, a necessary condition for a consistent AdS/MERA correspondence is that the MERA parameters k and χ satisfy the constraint (2.4.12). Importantly, this implies that, for a well-defined semiclassical spacetime (for which $c \gg 1$), the

bond dimension χ must be exponentially large in the size of the AdS scale compared to the Planck scale.

Let us also note that we can still obtain a bound from Eq. (2.4.11), albeit a weaker one, without using the result of Eq. (2.4.10). Recall that this latter result relies on having unentangled ancillae in the MERA. This is not necessarily the case for other tensor network bulk constructions, as we will subsequently discuss. As such, if we disregard the result of Eq. (2.4.10), we still have by virtue of their definitions that $f(k) \leq k - 1$ and $\eta \leq 1$. The following weaker but more general bound on the central charge therefore follows from Eq. (2.4.11) for such generalized tensor networks:

$$c \leq 12 \left(\frac{k-1}{\ln k} \right) \ln \chi. \tag{2.4.13}$$

2.5   Constraints from Bulk Entanglement Entropy

In addition to the compatibility conditions from geodesic matching and boundary entanglement entropy, it is well-motivated to seek out any other possible quantities that can be computed in both the MERA and AdS/CFT frameworks, so as to place further constraints on any AdS/MERA correspondence. One important example of such a quantity is the entropy associated with regions in the bulk, as opposed to on the boundary.

2.5.1   The Bousso bound

The notion of placing bounds on the entropy of regions of spacetime in a quantum gravity theory has been explored for many years, first in the context of black hole thermodynamics [45] and the Bekenstein bound [46] and later in more general holographic contexts, culminating in the covariant entropy bound, *i.e.*, the Bousso bound [47, 48].

The statement of the Bousso bound is the following: given a spacelike surface $\mathcal{B}$ of area A, draw the orthogonal null congruence on the surface and choose a direction in which the null generators have non-positive expansion. Let the null geodesics terminate at caustics, singularities, or whenever the expansion becomes positive. The null hypersurface swept out by these null geodesics is called the *lightsheet*. Then the entropy S going through the lightsheet is less than $A/4G$.

Let our spacelike surface $\mathcal{B}$ be a 2-ball of area A on a spacelike slice of AdS and choose as the lightsheet the ingoing future-directed null congruence. This lightsheet

will sweep out the entire interior of $\mathcal{B}$ and will terminate at a caustic at the center of $\mathcal{B}$. Since the system is static, the entropy S passing through this lightsheet is the entropy of the system on $\mathcal{B}$, which by the Bousso bound satisfies

$$S(\mathcal{B}) \leq \frac{A}{4G}. \tag{2.5.1}$$

It is natural to cast the Bousso bound as a constraint on the dimension of the bulk Hilbert space. As argued in Ref. [49], the thermodynamic entropy of a system about which we only know the boundary area A is just the logarithm of the dimension of the true Hilbert space of the bulk region in question (as opposed to the naïve Hilbert space in quantum field theory), which the Bousso bound implies is less than $A/4G$.[1] As such, if we denote the Hilbert space of $\mathcal{B}$ by $\mathcal{H}_{\mathcal{B}}$, let us replace Eq. (2.5.1) with the slightly more concrete statement

$$\ln \dim \mathcal{H}_{\mathcal{B}} \leq \frac{A}{4G}. \tag{2.5.2}$$

2.5.2 A MERA version of the Bousso bound

Our aim is to compute both sides of the inequality (2.5.2) using the MERA. For this calculation, it is instructive to change our parametrization of the hyperbolic plane from coordinates (x, z), which take values in the half-plane $z > 0$, to coordinates (ρ, θ), which take values in a disk $0 \leq \rho < 1$, $0 \leq \theta < 2\pi$. Embeddings of the MERA in a disk are often depicted in the literature, *e.g.*, [53]; here we make this coordinate transformation explicit, since we wish to carefully study the geometric properties of the MERA.

To begin, consider a MERA consisting of a single tree that contains a finite number of layers m. This situation is illustrated in Fig. 2.6a for $k = 2$ and $m = 4$. Note that such a MERA begins with a top-level tensor at the mth level that seeds the rest of the MERA in the IR.

The base of the MERA is made up of k^m sites. Without loss of generality, let us locate the leftmost site of the base of the MERA at $x = 0$, so that the UV-most sites sit at coordinates $(x, z) = (na, a)$, where $n = 0, 1, 2, \ldots, (k^m - 1)$ as shown

---

Moreover, it is known that there exists an asymptotically-AdS bulk configuration that saturates the Bousso bound, namely, the BTZ black hole [50, 51], which further implies that $\ln \dim \mathcal{H}_{\mathcal{B}}$ in fact equals $A/4G$. However, we will not need this stronger assertion in what follows. A similar but unrelated result equating the area of a region with its entanglement entropy in vacuum was obtained in Ref. [52].
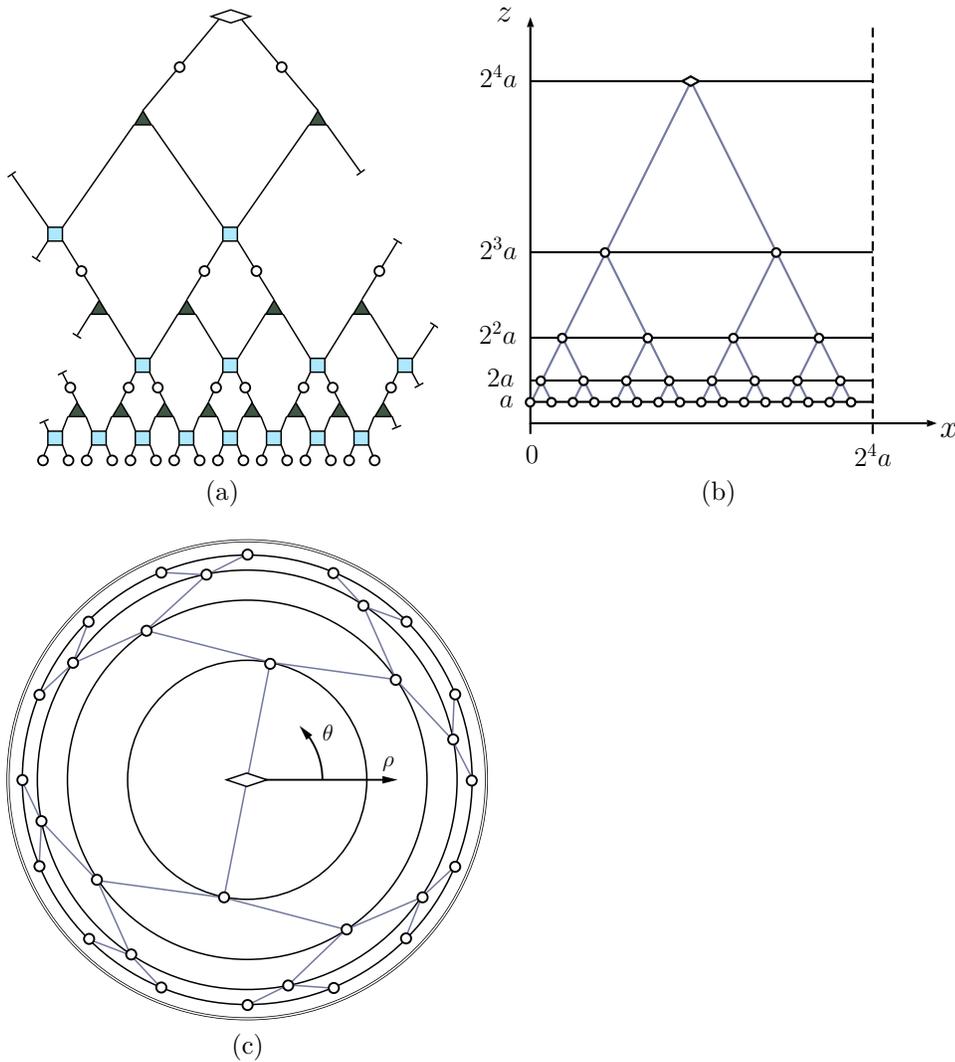
Figure 2.6. (a) A $k = 2$ MERA consisting of $m = 4$ layers and with periodic boundary conditions, (b) the corresponding embedding in (x, z) coordinates, and (c) the embedding in (ρ, θ) coordinates.

in Fig. 2.6b. Let us also assume periodic boundary conditions for this MERA and hence identify $x = 0$ and $x = k^m a$.

Next, define the coordinates (ρ, θ) as follows:

$$\rho = \frac{k^m a - z}{k^m a},$$
$$\theta = 2\pi \frac{x}{k^m a}. \tag{2.5.3}$$

In these coordinates, the metric reads

$$ds^2 = \frac{L^2}{(1-\rho)^2} \left[d\rho^2 + \left(\frac{d\theta}{2\pi}\right)^2 \right], \tag{2.5.4}$$

cf. Eq. (2.3.1). This embedding of the MERA is shown in Fig. 2.6c; the top-level tensor always sits at $\rho = 0$ and the lower layers of the MERA are equally spaced on circles of radii $1/2$, $3/4$, $7/8$, ... that are centered at $\rho = 0$.

More generally, one could construct a top-level tensor that has T legs, each of which begets a tree of sites. In this case, $x = 0$ and $x = Tk^{m-1}a$ are identified, so one should define the angular variable as $\theta \equiv 2\pi x/(Tk^{m-1}a)$. The metric (2.5.4) is correspondingly modified and reads

$$ds^2 = \frac{L^2}{(1-\rho)^2} \left[d\rho^2 + \frac{T^2}{k^2}\left(\frac{d\theta}{2\pi}\right)^2 \right]. \tag{2.5.5}$$

This situation is depicted in Fig. 2.7. (If $T = k$, however, then it is not necessary to introduce any new structure in addition to the disentanglers and isometries that were already discussed, *i.e.*, one may take the top-level tensor to be one of the isometries.)

We may immediately compute the right-hand side of Eq. (2.5.2). Let the ball $\mathcal{B}$ be centered about $\rho = 0$, and suppose $\mathcal{B}$ contains the top-level tensor, the sites at the top tensor's legs, and then the first $N_\mathcal{B}$ generations of the MERA emanating from these sites, as indicated in Fig. 2.7. The boundary of $\mathcal{B}$ is a circle at constant ρ, so its circumference according to the MERA is $A = Tk^{N_\mathcal{B}}L$. As such, we may write

$$\frac{A}{4G} = \frac{Tk^{N_\mathcal{B}}L}{4G} = \frac{Tk^{N_\mathcal{B}}c}{6}, \tag{2.5.6}$$

where in the second equality we used the Brown-Henneaux relation, Eq. (2.4.2).

How one evaluates the left-hand side of Eq. (2.5.2) using the MERA is not as immediate. Recall that $\mathcal{H}_\mathcal{B}$ is the Hilbert space of *bulk states*. The MERA, however, does not directly prescribe the quantum-gravitational state in the bulk; it is not
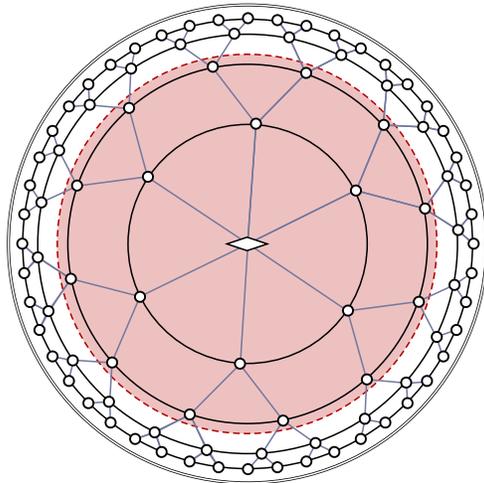
Figure 2.7. Disk parametrization of the Poincaré patch of AdS in which a MERA has been embedded. The top tensor of the MERA shown has $T = 6$. The shaded region is a ball $\mathcal{B}$, which is this case contains $N_\mathcal{B} = 1$ generation.

by itself a bulk-boundary dictionary. As we mentioned in Sec. 2.2.2, the minimal assumption that one can make is to posit the existence of a bulk Hilbert space factor V_{bulk} associated with each MERA site that is not located at the top tensor. To keep the assignment general, we assign a factor V_T to the top tensor. The dimensionality of each V_{bulk} factor should be the same in order to be consistent with the symmetries of the hyperbolic plane. The assumption of a Hilbert space factor at every MERA site is minimal in the sense that it introduces no new structure into the MERA. A true AdS/MERA correspondence should dictate how states in the bulk Hilbert space are related to boundary states. However, for our analysis, it is enough to simply postulate the existence of the bulk Hilbert space factors V_{bulk} and V_T, each of which may be thought of as localized to an AdS-scale patch corresponding to the associated MERA site.

In addition to the site at the top tensor, the number of regular MERA sites that the ball $\mathcal{B}$ contains is given by

$$\mathcal{N}_\mathcal{B} = T \sum_{i=0}^{N_\mathcal{B}} k^i = T \left(\frac{k^{N_\mathcal{B}+1} - 1}{k - 1} \right). \tag{2.5.7}$$

As such, the Hilbert space of bulk states restricted to $\mathcal{B}$ is $\mathcal{H}_\mathcal{B} = (V_{\text{bulk}})^{\otimes \mathcal{N}_\mathcal{B}} \otimes V_\text{T}$. Next, suppose that $\dim V_{\text{bulk}} = \tilde{\chi}$ and that $\dim V_\text{T} = \tilde{\chi}_\text{T}$, where, like χ, $\tilde{\chi}$ and $\tilde{\chi}_\text{T}$

are some fixed, $N_\mathcal{B}$-independent numbers. Then $\dim \mathcal{H}_\mathcal{B} = \tilde{\chi}_\mathrm{T}(\tilde{\chi}^{N_\mathcal{B}})$. Note that one would expect χ and $\tilde{\chi}$ to have a very specific relationship in a true bulk/boundary correspondence, the nature of which will be explored later in this section. Combining Eqs. (2.5.6) and (2.5.7), the dimensionality of $\mathcal{H}_\mathcal{B}$ is upper bounded as follows:

$$\ln \dim \mathcal{H}_\mathcal{B} \leq \frac{A}{4G} \quad \implies \quad T\left(\frac{k^{N_\mathcal{B}+1}-1}{k-1}\right)\ln \tilde{\chi} + \ln \tilde{\chi}_\mathrm{T} \leq \frac{Tk_\mathcal{B}^N c}{6}. \qquad (2.5.8)$$

After isolating c in Eq. (2.5.8) and using the result of Eq. (2.4.11), we find that

$$c = 12\eta f(k)\frac{\ln \chi}{\ln k} \geq 6\left(\frac{k^{N_\mathcal{B}+1}-1}{k^{N_\mathcal{B}}(k-1)}\ln \tilde{\chi} + \frac{1}{Tk^{N_\mathcal{B}}}\ln \tilde{\chi}_\mathrm{T}\right). \qquad (2.5.9)$$

Next, let us consider this inequality in the limit of large $N_\mathcal{B}$. A motivation for this limit is the fact that the natural scale of validity of an AdS/MERA correspondence is super-AdS, as was established in Sec. 2.3. Moreover, by virtue of its definition, there is always an ambiguity of order the AdS scale in the radius of the ball $\mathcal{B}$. That is, the region in AdS denoted by $\mathcal{B}$ is only well-defined in the MERA if $\mathcal{B}$ is large compared to the AdS scale L. Taking the limit of large $N_\mathcal{B}$, Eq. (2.5.9) reduces to

$$\eta f(k) \geq \frac{k\ln k}{2(k-1)}\left(\frac{\ln \tilde{\chi}}{\ln \chi}\right). \qquad (2.5.10)$$

By using the bound on $\eta f(k)$ given by Eq. (2.4.10), we arrive at a constraint on k, χ, and $\tilde{\chi}$:

$$\frac{k^2\ln k}{2(k-1)^2}\left(\frac{\ln \tilde{\chi}}{\ln \chi}\right) \leq 1. \qquad (2.5.11)$$

In principle, the above inequality could be satisfied for any k, provided that the dimension $\tilde{\chi}$ of the factors V_bulk can be arbitrarily chosen with respect to the bond dimension χ. However, the essence of holography, that the bulk and boundary are dual descriptions of the same degrees of freedom and therefore have isomorphic Hilbert spaces [14], implies a relation between χ and $\tilde{\chi}$. Namely, for a MERA with a total of N levels of sites in the bulk strictly between the UV-most level and the top-level tensor, the number of bulk sites $\mathcal{N}_\mathrm{bulk}$ that are not located at the top tensor is given by Eq. (2.5.7) with $N_\mathcal{B} = N$, and the number of sites in the boundary description is $\mathcal{N}_\mathrm{boundary} \equiv Tk^{N+1}$. The bulk Hilbert space thus has dimension $\tilde{\chi}^{\mathcal{N}_\mathrm{bulk}}\tilde{\chi}_\mathrm{T}$ and the

boundary Hilbert space has dimension $\chi^{\mathcal{N}_{\text{boundary}}}$. Equating[2] the dimension of the bulk and boundary Hilbert spaces then yields

$$\frac{\ln \tilde{\chi}}{\ln \chi} = \frac{1}{\mathcal{N}_{\text{bulk}}} \left(T k^{N+1} - \frac{\ln \tilde{\chi}_{\text{T}}}{\ln \chi} \right) \overset{N \, \text{large}}{\to} k - 1, \qquad (2.5.12)$$

where we took the limit of N large, consistent with Eq. (2.5.10) and in keeping with the expectation that the UV cutoff be parametrically close to the boundary at $\rho = 1$. Putting together Eqs. (2.5.11) and (2.5.12), we obtain a constraint on k alone:

$$\frac{k^2 \ln k}{2(k-1)} \leq 1. \qquad (2.5.13)$$

This constraint cannot be satisfied for any allowed value of the rescaling factor k, which must be an integer greater than or equal to 2. We thus learn that a conventional MERA cannot yield a consistent AdS/MERA correspondence. The MERA cannot simultaneously reproduce AdS geodesics, respect the Ryu–Takayanagi relation, and (using the only construction for the bulk Hilbert space available to the MERA by itself) satisfy the Bousso bound. That is, there exists no choice of MERA parameters that can faithfully reproduce geometry, holographic properties, and bulk physics.

If we relax this bound and, instead of Eq. (2.4.10), only observe the weaker, natural bounds $\eta \leq 1$ and $f(k) \leq k - 1$ as discussed at the end of Sec. 2.4.3, the constraint (2.5.13) is correspondingly modified:

$$\frac{k \ln k}{2(k-1)} \leq 1. \qquad (2.5.14)$$

In contrast to Eq. (2.5.13), this latter bound can be satisfied, but only for $k = 2$, 3, or 4. As such, other AdS/tensor network correspondences, in which the ancillae are perhaps entangled and therefore do not describe a conventional MERA, are not ruled out. Note that we never needed to compute bulk entanglement entropy explicitly — and therefore did not need to treat separately the possibility of entanglement among ancillae — because we cast the Bousso bound as a constraint on the size of the bulk Hilbert space itself. The appearance of η in Eq. (2.5.10) corresponds to entanglement

---

[2]We recognize that there are other proposals [27, 54] that do not require an exact equivalence between the bulk and boundary Hilbert spaces, but, even in these cases, there is the requirement of an exact equivalence between the logical qubits on the boundary with the Hilbert space of the bulk.

in the boundary theory as computed by the tensor network; Eqs. (2.5.10) and (2.5.12) still apply.

2.6 Conclusion

The notion of emergence of spacetime based on a correspondence between AdS and a tensor network akin to AdS/CFT is a tantalizing one. A necessary step in such a program is the evaluation and comparison of calculable quantities on both sides of the duality. In this work, we have subjected the proposed AdS/MERA correspondence to such scrutiny. To summarize, let us restate our three main findings:

1. In matching the discrete graph geometry of the MERA to the continuous geometry of a spatial slice of AdS, we demonstrated that the MERA describes geometry only on scales larger than the AdS radius. Concretely, as shown in Sec. 2.3, the proper length assigned to the spacing between adjacent sites in the MERA lattice must be the AdS scale.

2. By requiring that the entropy of a set of boundary sites in the MERA — whose computation is a discrete realization of the Ryu–Takayanagi formula — be equal to the CFT ground state entropy of the same boundary region in the thermodynamic limit, we obtained a constraint on the parameters that describe a MERA in terms of the CFT central charge [Eqs. (2.4.12) and (2.4.13)], which implies that the bond dimension χ must be exponentially large in the ratio of the AdS scale to the Planck scale.

3. In the natural construction of a bulk Hilbert space ($\mathcal{H}_{\text{bulk}}$) using the MERA, we used the Bousso bound to constrain the dimension of $\mathcal{H}_{\text{bulk}}$. When combined with our previous results, we found that any strict AdS/MERA correspondence cannot satisfy the resulting constraint, Eq. (2.5.13). Upon relaxing the definition of the MERA or allowing for additional structure, however, we obtained a looser constraint, Eq. (2.5.14), which may not rule out some other AdS/tensor network correspondences.

In particular, more general correspondences between AdS and MERA-like tensor networks, in which we allow the ancillae to be entangled when reproducing the CFT ground state [and for which Eq. (2.5.14) applies in place of Eq. (2.5.13)] are not ruled out by our bounds, provided that the rescaling factor $k = 2$, 3, or 4. Further, it is interesting to note that our bounds extend to states other than the vacuum that are

also described by a MERA. One such example, namely, states at finite temperature dual to black holes in AdS, is discussed in App. 2.B below.

While the consistency conditions that we found are specific to the MERA tensor network, many of the ideas and techniques that we used apply equally well to other tensor networks. In the EHM, for instance, the type of bulk Hilbert space dimensionality arguments that we made based on the covariant entropy bound may be directly transferred to the EHM. The same stringent final constraints that we derived do not apply to the EHM, however, since it is unclear to what extent the EHM reproduces the Ryu–Takayanagi formula (which renders the results of Sec. 5.2 inapplicable). Our bulk Hilbert space arguments similarly apply to the holographic error-correcting code proposal in Ref. [27], which furthermore purports to reproduce a version of the Ryu–Takayanagi formula. It is presently unknown, however, whether the boundary state of a holographic code can represent the ground state of a CFT, so an identification of entropies similar to the identification $S_{\mathrm{MERA}} = S_{\mathrm{CFT}}$, upon which our boundary entropy constraints so crucially depend, cannot yet be made.

In closing, we have found several consistency conditions that any AdS/MERA correspondence must satisfy. The totality of these constraints rules out the most straightforward construal of an AdS/MERA correspondence. Other interesting holographic correspondences that are described by tensor networks more general than the MERA and that respect all of our bounds may indeed be possible. Our consistency conditions are nice validity checks for these correspondences when applicable and in other cases they may inspire similar consistency conditions. The program of identifying the emergence of spacetime from the building blocks of quantum information is an ambitious one; stringent consistency conditions, such as those presented in this paper, are important for elucidating the subtleties in this quest and in providing guidance along the way.

2.A Entropy Bound for General MERAs

Following the method presented in Ref. [37], let us compute an upper bound for the entanglement entropy of a region B consisting of ℓ_0 sites in a MERA with rescaling factor k. We will use the notation of Ref. [37] throughout.

First, recall the result from Ref. [37] that the entanglement entropy of a region consisting of ℓ_0 sites is bounded by

$$S_{\mathrm{MERA}}(\ell_0; B) \leq (\ell_{m'} + N_{m'}^{\mathrm{tr}}) \ln \chi. \tag{2.A.1}$$

The quantity $\ell_{m'}$ is the width of the causal cone at depth m' and $N_{m'}^{\mathrm{tr}} = \sum_{m=0}^{m'-1} n_m^{\mathrm{tr}}$ is the total number of sites that are traced out along the boundary of the causal cone. In other words, $N_{m'}^{\mathrm{tr}}$ is the number of bonds that are cut by the causal cone up to a depth m' (*cf.* Fig. 2.4). The quantity $\ln \chi$ is the maximum entanglement entropy that each site that is traced out could contribute to $S_{\mathrm{MERA}}(\ell_0; B)$. Note that Eq. (2.A.1) holds for all $m' \geq 0$.

The width of the causal cone for a given m' depends sensitively on the structure of the MERA. In particular, the number of sites that are traced out at each renormalization step depends on the choice of disentanglers, as well as how they are connected to the isometries. For instance, in a MERA with a rescaling factor k, any given disentangler could have anywhere from 2 up to k incoming and outgoing legs. (It should be reasonable to require that any disentangler can have no more than k incoming and k outgoing legs so that it straddles no more than two isometries.) It is thus clear that the number of bonds that one cuts when drawing a causal cone, and hence the entanglement entropy of the region subtended by that causal cone, depends on the choice of disentanglers and connectivity.

Nevertheless, we can compute an upper bound for $S_{\mathrm{MERA}}(\ell_0; B)$ by considering a worst-case scenario for the number of bonds cut by the causal cone. We begin by asking: What is the largest number of bonds that a causal cone could cut in one renormalization step at a depth m'? The layout of disentanglers and isometries that produces this situation is shown at one side of a causal cone in Fig. 2.8. If the causal cone at the bottom of the renormalization step incorporates a single bond that goes into a disentangler accepting k bonds, then the causal cone must cut the other $k - 1$ bonds entering the disentangler. Then if this disentangler is arranged

so that its leftmost outgoing bond is the first bond to enter an isometry from the right, the causal cone must cut the other $k - 1$ bonds entering the isometry. If this arrangement is mirrored on the other side of the causal cone, we see that $4(k - 1)$ bonds are cut by the causal cone in this renormalization step, *i.e.*, $n_{m'}^{\text{tr}} = 4(k - 1)$.
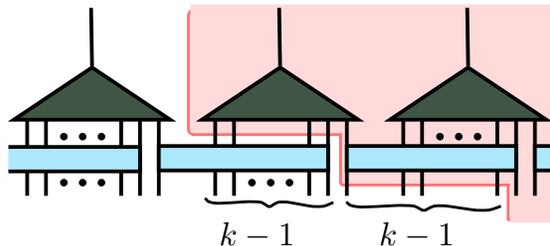
Figure 2.8. Left side of a causal cone that cuts the maximum possible number of bonds over the course of one renormalization step. The rectangles are disentanglers that accept k bonds as input and the triangles are isometries that coarse-grain k bonds into one. The causal cone is the shaded region. If this situation is mirrored on the right side of the causal cone, then $4(k - 1)$ bonds are cut in this renormalization step.

Recall that for any finite ℓ_0, after a fixed number of renormalization steps, the width of the causal cone remains constant for any further coarse-grainings. The depth at which this occurs is called the crossover scale and is denoted by $\bar{m}$. Therefore, the causal cone will cut the largest possible number of bonds when the arrangement described above and depicted in Fig. 2.8 occurs at every step up until the crossover scale. Then, by Eq. (2.A.1), the entropy bound is given by

$$S_{\text{MERA}}(\ell_0; B) \leq (\ell_{\bar{m}} + 4(k - 1)\bar{m}) \ln \chi, \qquad (2.A.2)$$

where $\ell_{\bar{m}}$ is the width of the causal cone at the crossover scale.

For any given causal cone in a MERA with scale factor $k \geq 2$, the maximum number of additional sites the causal cone can pick up at some level m' is $4(k - 1)$. Therefore, for a causal cone that contains $\ell_{m'}$ sites at depth m', the number of sites in the causal cone after one renormalization step $\ell_{m'+1} \leq \lceil (\ell_{m'} + 4(k - 1))/k \rceil \leq \ell_{m'}/k + 5$. Applying the relation recursively, we find that the number of sites $\ell_{m'}$ at any layer $m' < \bar{m}$ is bounded,

$$\ell_{m'} \ \leq \ \frac{\ell_0}{k^{m'}} + 5 \sum_{m=1}^{m'} \frac{1}{k^m} \ \leq \ \frac{\ell_0}{k^{m'}} + 5 \,. \qquad (2.A.3)$$

Setting $m' = \bar{m}$, it trivially follows that the crossover scale obeys $\bar{m} \leq \log_k \ell_0$. Furthermore, we notice that this is the scale at which the entanglement entropy is minimized if we trace over the remaining sites. In other words, the number of bonds cut by going deeper into the renormalization direction is no less than the bonds cut horizontally, so $4(k-1) \geq \ell_{\bar{m}}$ [3]. Applying the bounds for $\bar{m}$ and $\ell_{\bar{m}}$ on Eq. (2.A.2), we arrive at an upper bound on $S_{\text{MERA}}(\ell_0; B)$ for a k-to-one MERA,

$$S_{\text{MERA}}(\ell_0; B) \leq 4(k-1)(1 + \log_k \ell_0) \ln \chi. \qquad (2.A.4)$$

When ℓ_0 is parametrically large, we neglect the $\mathcal{O}(1)$ contribution to the bound on $S_{\text{MERA}}(\ell_0; B)$, which yields Eq. (2.4.5).

2.B BTZ Black Holes and Thermal States in AdS/MERA

Thus far, we have found constraints on the structure of a MERA that can describe CFT states dual to the AdS_3 vacuum. One might ask whether these results extend to other constructions that exist in three-dimensional gravity. Although pure gravity in AdS_3 has no local or propagating degrees of freedom, there exist interesting non-perturbative objects, namely, BTZ black holes [50]. In this appendix, we extend our constraints on boundary entanglement entropy to these objects.

The non-rotating, uncharged BTZ black hole solution is given in Schwarzschild coordinates by

$$ds^2 = -\frac{(r^2 - r_+^2)}{L^2}dt^2 + \frac{L^2}{(r^2 - r_+^2)}dr^2 + r^2 d\phi^2 \,, \qquad (2.B.1)$$

with a horizon at $r = r_+$. Noting that Euclidean time is compactified by identifying $\tau \sim \tau + 2\pi L^2 / r_+$, the horizon temperature of the black hole is given by $T = r_+/2\pi L^2$. Additionally, the Bekenstein–Hawking entropy of the black hole is

$$S_{\text{BH}} = \frac{\text{Area}}{4G} = \frac{\pi r_+}{2G} \,. \qquad (2.B.2)$$

Let us now consider applying a MERA with rescaling factor k and bond dimension χ to a CFT at a finite temperature, where instead of minimizing the energy of

---

Alternatively, we can see this from a heuristic argument by noting that the crossover scale is the scale at which the causal cone has a constant width for further coarse-grainings, *i.e.*, $(\ell_{\bar{m}} + 4(k-1))/k \approx \ell_{\bar{m}}$. Therefore, $\ell_{\bar{m}} \lesssim 4 \leq 4(k-1)$.

the boundary state, one minimizes the free energy. In the CFT, turning on a temper-
ature introduces a scale, going as the inverse temperature, which screens long-range
correlations. Thus, the state will have classical correlations in addition to entan-
glement and the effect of a finite temperature on the entanglement entropy is the
appearance of an extensive contribution. As one runs the MERA and coarse-grains,
the thermal correlations that cannot be removed become more relevant. The MERA,
which is unable to remove the extensive contribution, truncates at a level with mul-
tiple sites. The schematic entanglement renormalization process is illustrated in
Fig. 2.9. The state at the top level effectively factorizes, where each factor appears
maximally mixed [25, 36]. A tractable realization of this tensor network structure
recently appeared in Ref. [55], which found a MERA representation of a thermal
state.
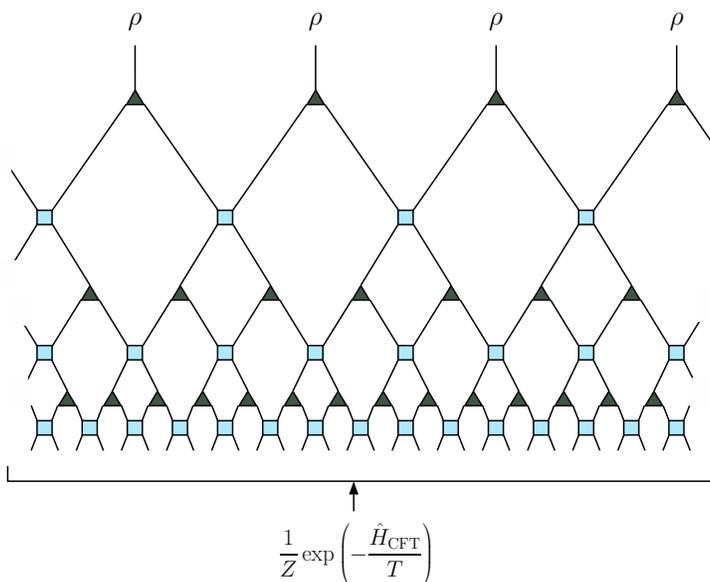
Figure 2.9. The MERA, when applied to a thermal CFT state $Z^{-1} \exp(-\hat{H}_{\mathrm{CFT}}/T)$, where
$Z = \mathrm{tr}(\exp(-\hat{H}_{\mathrm{CFT}}/T))$, truncates after a finite number of layers. The boundary state at
the top of the truncated MERA effectively factorizes into a product of maximally mixed
states $\rho = I/\chi$.

Keeping in mind that the holographic dual of a finite-temperature state in the
CFT is a black hole in AdS, where the temperature of the CFT corresponds to
the Hawking temperature of the black hole, we note that the truncated MERA
is suggestive of a black hole horizon [25]. If the MERA is to be interpreted as a

discretization of the geometry, then the geometry has ended at some scale. Also, as we approach the horizon, the amount of Hawking radiation that we see increases and the temperature measured by an observer at the horizon diverges. The density matrix of some system in the infinite-temperature limit is given by the product of a maximally mixed state at each site, just like the state at the top of the MERA. It is important to note that, as was pointed out in Ref. [55], in order to reproduce the correct thermal spectrum of eigenvalues, a small amount of entanglement must be present between the sites at the horizon. If the bond dimension were taken to be infinite, then the sites at the horizon truly would factorize. But for a finite bond dimension, one should really think of the horizon as a high-temperature state, with sites effectively factorized.

For small regions on the boundary, the length of the subtending bulk geodesic is subextensive and so the Ryu–Takayanagi formula maintains that the boundary region's entanglement entropy is subextensive as well. However, if we consider a large enough region on the boundary, the geodesic will begin to probe the horizon of the black hole. The geodesic will run along the black hole horizon and pick up an extensive contribution to the entropy. We consider a boundary theory living on a lattice consisting of n_b sites, with total system coordinate length $x_{sys} = n_b a$. In the limit as r approaches the boundary in the metric (2.B.1), we see that $T x_{sys} = r_+/L$, as was pointed out in Refs. [21, 44]. We further note that this implies that the system coordinate size is of order AdS radius, $x_{sys} = 2\pi L$.

Let us now view the MERA of Fig. 2.9 as a discretization of a BTZ spacetime and repeat the analysis of Sec. 2.3. In this discretization, the layers of the MERA lie along circles of fixed radius r in the coordinates of Eq. (2.B.1). Again, we ask what proper length L_1 separates sites in any given layer of the MERA.

First, note that a path at fixed r_0 that subtends an angle ϕ_0 has proper length $r_0 \phi_0$. At the boundary of the MERA, we consider a region defined by $0 \leq \phi \leq \phi_0 = 2\pi x_0/x_{sys}$, where x_0 is the coordinate length of the interval, consisting of ℓ_0 lattice sites. The boundary of the MERA is at a fixed radius $r = r_b$. Naturally, the boundary radius r_b can be interpreted as a UV cutoff and is related to the lattice spacing a by $r_b = L^2/a$ [21]. By equating the proper distance of the region in the MERA, $\ell_0 L_1$, with that at the boundary of the BTZ spacetime, $r_b \phi_0$, we find the proper length between horizontal bonds to be $L_1 = L$.

With the foresight that the top of the MERA is suggestive of a black hole horizon with proper length $2\pi r_+$, the number of sites at the final layer is therefore $n_h = 2\pi r_+/L$. This further tells us that the MERA truncates after a finite number of

layers m, given by

$$m = \log_k \left(\frac{n_{\mathrm{b}}}{n_{\mathrm{h}}} \right) = \log_k \frac{1}{2\pi T a} \,. \tag{2.B.3}$$

This coincides with the conclusion in Refs. [55, 56] that the MERA representation of a thermal state is obtained after $\mathcal{O}(\log_k(1/T))$ iterations of coarse-graining.

Now consider a region B on the boundary consisting of ℓ_0 sites and for which the corresponding geodesic contains a segment running along the BTZ horizon. The subextensive contribution to the entropy in the MERA is exactly as before, in which we pick up at most $\ln \chi$ from each bond we cut with the causal cone of the region B. Furthermore, we will now pick up an extensive contribution from the horizon, where the number of horizon sites within the causal cone is ℓ_{h} and each such site in the product state on the horizon contributes maximally to the entropy by an amount $\ln \chi$. Combining the contributions, we find

$$S_{\mathrm{MERA}}(B) = 4\eta_B f_B(k) \log_k \left(\frac{\ell_0}{\ell_{\mathrm{h}}} \right) \ln \chi + \ell_{\mathrm{h}} \ln \chi \,. \tag{2.B.4}$$

Recall that the entanglement entropy of a single interval B of coordinate length x_0 in a CFT at finite temperature [42] is given, up to a non-universal constant, by

$$S_{\mathrm{CFT}}(B) = \frac{c}{3} \ln \left(\frac{1}{\pi a T} \sinh \pi x_0 T \right) \,, \tag{2.B.5}$$

where x_0 is much smaller than the total system size x_{sys}. The standard field-theoretic derivation of the above entropy is done by computing the Euclidean path integral on an n-sheeted Riemann surface and analytically continuing to find the von Neumann entropy. The same result can be derived by computing geodesic lengths on spatial slices of BTZ spacetimes and making use of the Ryu–Takayanagi formula.

When $T \to 0$ in Eq. (2.B.5), we recover the usual result (2.4.1). In the $T \to \infty$ limit, the von Neumann entropy gives the usual thermal entropy as entanglement vanishes. Taking $Tx_0 \gg 1$, the leading and subleading contributions to the entanglement entropy are

$$S_{\mathrm{CFT}} = \frac{c}{3} \pi x_0 T + \frac{c}{3} \ln \frac{1}{2\pi a T} \,, \tag{2.B.6}$$

where the first term is the thermal entropy for the region B.

Now let us consider a finite-temperature CFT that is dual to a BTZ black hole with horizon temperature $T = r_+/2\pi L^2$. In terms of geometric MERA parameters,

we find that Eq. (2.B.6) becomes

$$S_{\text{CFT}} = \frac{c}{6}\,\ell_{\text{h}} + \frac{c}{3}m\ln k\,. \tag{2.B.7}$$

Here we used the fact that $\ell_{\text{h}} = x_0 r_+/L^2$ as well as Eq. (2.B.3), where we note that m can also be written as $\log_k(\ell_{\text{b}}/\ell_{\text{h}})$. The result (2.B.7) coincides precisely with the extensive and subextensive contributions calculated using the MERA in Eq. (2.B.4) provided that $c/\ln\chi \sim \mathcal{O}(1)$. Therefore, we find that the truncated MERA correctly captures the entanglement structure of thermal CFT states and their dual BTZ spacetimes. These conclusions are in agreement with those in [35, 56].

As a check of the claim that c and $\ln\chi$ should be of the same order, we can compare the horizon entropy given by the contribution from the sites at the final layer with the Bekenstein–Hawking entropy (2.B.2) of a BTZ black hole. There are n_{h} sites comprising the horizon, each with Hilbert space dimension χ. The system is in the infinite-temperature limit — and hence described by a maximally mixed density matrix, with entropy contribution $\ln\chi$ from each site — so

$$S_{\text{horizon}} = n_{\text{h}}\ln\chi\,. \tag{2.B.8}$$

Making use of the Brown–Henneaux relation and requiring the entropy (2.B.8) to coincide with the Bekenstein-Hawking entropy, we again find that $c/\ln\chi \sim \mathcal{O}(1)$. More specifically, taking the counting to be precise, we find that

$$c/\ln\chi = 6\,, \tag{2.B.9}$$

which is qualitatively in agreement with the previous conclusion (2.4.12) that the Hilbert space dimension must be exponentially large in c.

With this relation, the extensive terms in Eqs. (2.B.4) and (2.B.7) agree precisely. Further identifying the subextensive terms, we find $\eta_B f_B(k) = (\ln k)/2$. If we then impose the constraint (2.4.10), we find that

$$\frac{k\ln k}{2(k-1)} \leq 1\,. \tag{2.B.10}$$

This last inequality exactly reproduces Eq. (2.5.14) and thus constrains k to be 2, 3, or 4. Interestingly, we have found the weaker of the two bounds derived in Sec. 3.3, without needing to consider the Bousso bound.

As desired, the truncated MERA computation of entanglement entropy agrees with the expected entanglement entropy given by the application of the Ryu–Takayanagi formula to the length of the minimal surface in a BTZ spacetime. The fact that the results of matching boundary entanglement entropy given in Sec. 5.2 further hold in BTZ spacetimes might not be too surprising given that such spacetimes are quotients of pure AdS_3.

Splitting Spacetime and Cloning Qubits: Linking No-Go Theorems across the ER=EPR Duality

3.1 Introduction

The connection between entanglement and geometry is an unexpected stepping-stone on the path to an understanding of quantum gravity. Historically originating from black hole thermodynamics [57, 58] and later in the context of the holographic principle [59, 60], the AdS/CFT correspondence [12–14], entropy bounds [61], and the Ryu–Takayanagi formula [15], the relation between quantum entanglement and spacetime geometry is increasingly thought to be an important feature of a consistent theory of quantum gravity. Underscoring this view is recent work on deriving the Einstein equations holographically from entanglement constraints [62] and perhaps even spacetime itself from qubits [63, 64]. However, significant puzzles remain. The classic black hole information paradox [65, 66] has given way to new questions about black hole interiors and their entanglement with Hawking radiation [67, 68]. One of the most drastic, albeit promising, proposals to arise from these debates is the so-called ER=EPR duality [16].

The ER=EPR correspondence [16] is a compelling [69, 70] proposal for an exact duality between Einstein-Podolsky-Rosen (EPR) pairs [71], that is, qubits entangled in a Bell state [72], and nontraversable wormholes, that is, Einstein-Rosen (ER) bridges [73–75]. More specifically, the ER=EPR proposal generalizes the notion of entangled black hole pairs at opposite ends of an ER bridge, by asserting that every pair of entangled qubits is connected by a Planck-scale quantum wormhole. The proposal, if true, would have profound implications for AdS/CFT and suggest a solution to the firewall paradox of Ref. [67], not to mention the fundamental shift it would induce in our understanding of both quantum mechanics and general relativity.

The ER=EPR correspondence might allow the exploration of gravitational analogues of fundamental properties of quantum systems (and vice versa). In particular, we can check whether there is a precise correspondence between no-go theorems in quantum mechanics and similar no-go theorems in gravity. Arguably the most celebrated no-go theorem in quantum mechanics is the no-cloning theorem [76], which prohibits the duplication of quantum states.

In this paper, we investigate the manifestation of the no-cloning theorem on the gravitational side of the ER=EPR duality. In particular, we show that violation of the no-cloning theorem is dual under ER=EPR to topology-changing processes in general relativity, which, via classical topology-conservation theorems [77–83], lead to causal anomalies through violation of the Hausdorff condition (which leads to the breakdown of strong causality), creation of closed timelike curves (CTCs), or

violation of the null energy condition (NEC) (which allows for wormhole traversability and hence CTCs). While the validity of ER=EPR requires both unitarity and wormhole nontraversability, it is interesting that these two requirements seem to be fundamentally related: the no-cloning theorem and the topology-conservation theorem, both of which are related to causality, are in fact dual no-go theorems under ER=EPR.

3.2   Quantum Cloning

Here, we reconstruct the standard argument for why the no-cloning theorem prohibits superluminal signaling [84]. Assume that cloning of states is allowed, that is, that there exists an operation that takes an arbitrary state $|\Psi\rangle$ in a product state with some $|0\rangle$ state and replaces the $|0\rangle$ state with $|\Psi\rangle$:

$$|\Psi\rangle_A |0\rangle_B \to |\Psi\rangle_A |\Psi\rangle_B. \tag{3.2.1}$$

Suppose that there exists an EPR spin pair, the state $(|00\rangle + |11\rangle)/\sqrt{2}$. We give one spin to each of a pair of individuals, Alice and Bob, who may then move to arbitrary spacelike separation. Alice now makes a decision as to the classical bit she wishes to communicate: to send a "1", she measures in the σ_z basis, while to send a "0", she does nothing.

Bob now proceeds to clone his qubit as in Eq. (3.2.1). Note that each of his cloned qubits remains maximally entangled with Alice's qubit, in violation of monogamy of entanglement, while remaining unentangled with each other. By measuring enough of his own qubits in the σ_z basis, Bob can determine, to any desired degree of confidence, whether Alice performed a measurement or not: his measurements will all yield the same result if Alice performed a measurement, but will be equally and randomly split between the two outcomes if she did not. As this experiment does not depend on their separation, Bob's utilization of cloning and their shared entanglement has allowed Alice to send one classical bit to Bob acausally.

3.3   Black Hole Cloning

In order to geometrically interpret the no-cloning theorem using the ER=EPR proposal, we need a system with both a high level of entanglement (like the EPR pair just considered) and a robust geometric description. One such system is the eternal

AdS-Schwarzchild black hole, which is described in AdS/CFT by two noninteracting large-N CFTs in a thermally entangled state on the boundary sphere [85, 86]:

$$|\Psi\rangle = \frac{1}{\sqrt{Z}} \sum_n e^{-\beta E_n/2} |n\rangle_L \otimes |n\rangle_R, \qquad (3.3.1)$$

where $|n\rangle_{L(R)}$ is the nth eigenstate on the left (respectively, right) CFT with energy E_n, β is the inverse temperature, and Z is the partition function. In this state, the reduced density matrices $\rho_{L,R}$ of either side are identically thermal. If both exterior regions of the geometry are considered [85–87], this state describes a spacetime consisting of two separate AdS-Schwarzchild regions that are spatially disconnected outside the horizon but linked by an ER bridge between a maximally entangled[1] pair of black holes with temperature β^{-1}. This is a concrete realization of ER=EPR: to reiterate, the two black holes are both maximally entangled (EPR) and connected by a nontraversable wormhole (ER). It will be convenient to consider the slight generalization of this setup in which the two black holes share the same asymptotic space. As discussed in Ref. [16], such black hole pairs can be naturally obtained as an instanton solution in a geometry with a constant magnetic field.

We now consider repeating the experiment in the previous section using entangled black holes instead of qubits, as depicted in Fig. 3.1. Alice and Bob, who live in an asymptotically-AdS spacetime, are each given access to a Schwarzschild black hole, with the two black holes, labeled A and B respectively, maximally entangled and therefore connected by a nontraversable wormhole. If Bob now clones all the degrees of freedom on his stretched horizon [88], he is left with two black holes B and B', each of which is connected by an ER bridge to Alice's black hole. That is, cloning is dual to change of spacetime topology under ER=EPR.

3.4 Changing Spacetime Topology

We now turn to the question of whether the double-wormhole geometry of Fig. 3.1 suffers from any inconsistencies in general relativity. Throughout, we assume that the Einstein equations hold and that the spacetime can be well described by a semiclassical geometry (which corresponds to a choice of how Bob implements the cloning).

The simplest interpretation of the geometry M in Fig. 3.2 is that, since horizon pairs AB and AB' are each in the thermofield double state (3.3.1), the geometries

---

[1]Strictly speaking, the state is only truly maximally entangled when $\rho_L = \rho_R = \mathbb{1}$, i.e., when $\beta \to 0$, but we adopt the terminological abuse of Ref. [16].
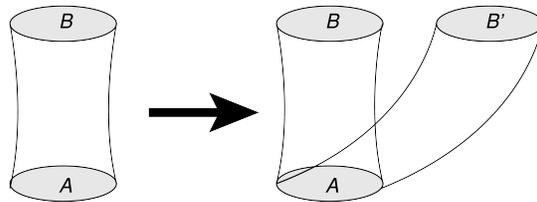
Figure 3.1. Illustration of the black hole cloning thought experiment in the context of the ER=EPR conjecture. If Bob has access to a device that can clone quantum states, he can transform black hole B, which is entangled with A, into two black holes B and B', each connected to A via an ER bridge.
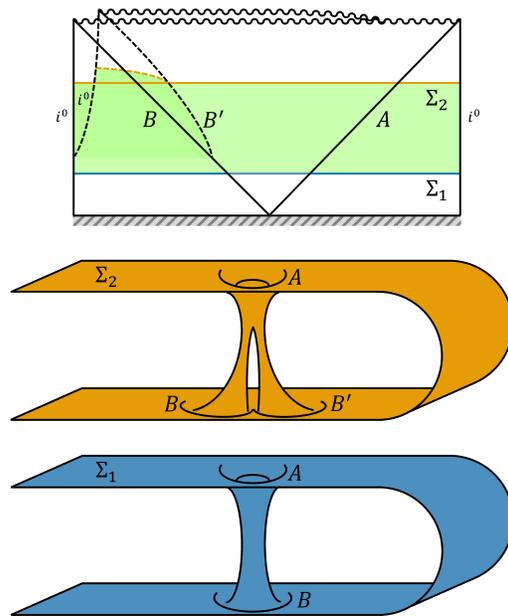
Figure 3.2. Penrose diagram for the topology-change process depicted in Fig. 3.1, with spatial slices Σ_1 (blue) and Σ_2 (orange) shown as embedding diagrams. The spacetime region M (green) is indicated; the compact region K with nontrivial topology is bounded by horizons A, B, and B'. All of the spatial infinities i^0 are identified, as the black holes share the same asymptotically-AdS spacetime. The diagonal stripes at the bottom of the Penrose diagram indicate that the half of the spacetime containing the past horizons is not shown.

of both wormholes are the same. In this case, the geometry after Bob performs the cloning simply consists of two separate sheets, each a copy of the original ER bridge, glued together along horizon A. Note that in this case M contains bifurcate

geodesics: any timelike geodesic intersecting horizon A after the cloning occurs will split into two timelike geodesics, one going along the sheet containing B and the other along the sheet containing B'. These timelike bifurcate curves indicate a breakdown of the Hausdorff condition, the requirement that for any two points $x \neq y$, there exist disjoint open sets $X \ni x$ and $Y \ni y$.[2] Since the bifurcate timelike curve in question has bounded (being a geodesic, zero) acceleration and moreover the non-Hausdorff boundary of M (horizon A) is codimension 1, it follows by a theorem of Håjíček [81] that M is not strongly causal. Strong causality is the requirement that for all points $p \in M$ there is an open neighborhood $P \ni p$ such that any timelike curve passing through P does so only once; this is a weaker condition than global hyperbolicity, so the setup depicted in Fig. 3.2 leads, via Håjíček's theorem, to breakdown of Cauchy evolution [90]. Intuitively, this happens because once a timelike curve intersects horizon A it becomes impossible to predict its future. If we wish to avoid immediately abandoning strong causality, we must relax the assumption that the geometry after cloning is merely a two-sheeted copy of the original ER bridge and instead turn to the question of whether the topology change induced by cloning is alone sufficient to guarantee a pathology for a spacetime that remains Hausdorff.

The topology change in question occurs in a localized region of spacetime. Let us define a partial Cauchy surface [79] to be a spacelike slice through the entire spacetime such that any causal (timelike or null) curve intersects the surface at most once. A 3-surface Σ is called externally Euclidean if there exists compact $\Gamma \subset \Sigma$ such that $\Sigma - \Gamma$ is diffeomorphic to Euclidean space minus a 3-ball, i.e., $\Sigma - \Gamma \simeq S^2 \otimes \mathbb{R}$. Given these definitions, we can draw two disjoint externally Euclidean partial Cauchy surfaces Σ_1 and Σ_2, where Σ_1 passes through horizons A and B before the cloning and Σ_2 passes through horizons A, B, and B' after the cloning, as shown in Fig. 3.2. Importantly, Σ_1 and Σ_2 are *not* diffeomorphic, $\Sigma_1 \not\simeq \Sigma_2$. Taking A, B, and B' to be centered on a line on Σ_2 and quotienting by the rotation group $SO(2)$ around this line, $\Sigma_1/SO(2)$ and $\Sigma_2/SO(2)$ are 2-manifolds with genera 1 and 2, respectively, and are therefore not topologically equivalent. The four-dimensional spacetime region whose boundary is $\Sigma_1 \cup \Sigma_2$, called M in Fig. 3.2, is externally Lorentzian: there exists a compact manifold K such that $M - K \simeq S^2 \otimes \mathbb{R} \otimes [0, 1]$, a timelike foliation of spacelike slices $S^2 \otimes \mathbb{R}$. Then Geroch's topology-conservation theorem [77–79] implies that, since $\Sigma_1 \not\simeq \Sigma_2$, M must contain a CTC.

While the existence of a CTC somewhere in spacetime is already problematic,

---

[2]Bifurcating geodesics imply failure of the Hausdorff condition, but the converse is not necessarily true; see, for example, the discussion of Taub-NUT space in Refs. [81, 89].

we can state a stronger result. We note that Σ_1 is a Cauchy surface for $M - K$, that is, for all $p \in M - K$, every future- and past-inextendible causal curve through p intersects Σ_1. Let us assume the generic condition, which asserts that every causal geodesic with tangent vector k^μ passes through some point for which

$$k^\alpha k^\beta k_{[\mu} R_{\nu]\alpha\beta[\rho} k_{\sigma]} \neq 0. \tag{3.4.1}$$

This means that every timelike or null geodesic experiences a tidal force at some point.[3] Then Tipler's topology-conservation theorem [79, 80] implies that since $\Sigma_1 \not\simeq \Sigma_2$, the NEC[4] must fail. That is, the topology change dual to cloning under ER=EPR implies that there must exist fields in the theory for which one can arrange an energy-momentum tensor $T_{\mu\nu}$ such that

$$T_{\mu\nu} k^\mu k^\nu < 0 \tag{3.4.2}$$

along some null vector k^μ.

Although violations of the NEC (see also Ref. [91]) have been shown to occur at a quantum level [92], it has not been shown that such violation is sufficient to allow unusual semiclassical gravitational behavior [16, 93]. However, the NEC violation in the present thought experiment implies macroscopic topology change that results from Bob's cloning procedure with, for example, astrophysical-scale entangled black holes. We conclude that violation of the no-cloning theorem is dual under ER=EPR to topology change and problems with causality, leading to CTCs (by Geroch's theorem) or strong violation of the NEC (by Tipler's theorem).

It is worth noting that the topology theorems do not rule out sensible processes like black hole pair production in the context of ER=EPR. If we consider entanglement as a conserved quantity [94], then creation of a pair of entangled black holes does not change the topology, as the ER bridge between them is formed in ER=EPR from the Planckian wormholes connecting the entangled vacuum. Moreover, the process of black hole pair creation is not well described semiclassically, so our results do not apply in that case; in contrast, the cloning process examined in this work can be treated in the setting of semiclassical geometry. Unlike pair production, cloning

---

[3]If the spacetime under consideration has some special symmetry allowing Eq. (3.4.1) to fail for some geodesic, we can enforce the generic condition by simply adding gravitational waves (that is, nonzero Weyl tensor) sufficiently weak to avoid nonnegligible back-reaction on the rest of our argument.

[4]While Ref. [79] states the theorem in terms of the weak energy condition, this can be strengthened to the NEC as stated in Ref. [80].

does violate the axioms of the topology-conservation theorems precisely because it involves non-unitarily creating entanglement (and therefore wormholes) that did not previously exist.

3.5 Wormholes and Causality

We have shown that violation of the no-cloning theorem is dual under ER=EPR either to immediate breakdown of Cauchy evolution or to severe violation of the NEC [Eq. (3.4.2)]. The latter implies the condition that allows for stabilization of wormholes; specifically, one must have violation of the *averaged* NEC [93, 95]. That is, a traversable ER bridge requires

$$\int_0^\infty T_{\mu\nu}k^\mu k^\nu \mathrm{d}\lambda < 0 \tag{3.5.1}$$

for some null geodesics with affine parameter λ and tangent vector k^μ. Ref. [93] exhibits a construction of a traversable ER bridge that just satisfies Eq. (3.5.1) within the wormhole while retaining nonnegative total energy.

The connection between wormhole stabilization and the NEC is highly relevant in the context of the ER=EPR correspondence, as the argument in Ref. [16] regarding the impossibility of using wormholes (and by duality, entanglement) to transmit information is critically dependent on the ER bridges pinching off too quickly to allow for signal traversal [75]; a stabilized wormhole would falsify this line of reasoning. Said another way, violation of the NEC plus the existence of wormholes leads to traversable wormholes, which would lead to causality violation. In particular, given a traversable ER bridge, one can immediately form a causal paradox (i.e., a closed signal trajectory) by simply moving the wormhole mouths far apart and giving them a small relative boost [93, 96]. The connection between topology change and causality violation in the gravitational sector is now explicit and is satisfyingly analogous to the connection between unitarity/no-cloning and causality on the quantum mechanical side of the ER=EPR duality.

3.6 Perspectives for Future Work

As we have seen, spacetime topology change leads inexorably to violation of causality, via either breakdown of the Hausdorff condition or creation of traversable wormholes. Using ER=EPR to translate this result to quantum mechanics, we find that violation

of the axioms of the topology-conservation theorems is dual to violation of monogamy of entanglement (i.e., cloning) and the existence of wormholes is dual to the existence of entanglement entropy. The logical flow of our reasoning is:

$$
\begin{array}{ccccc}
\text{C} & \& & \exists\,\text{QE} & \implies & \text{SLS} \\
\updownarrow & & \updownarrow & & \updownarrow \\
\Delta\text{T} & \& & \exists\,\text{WH} & \stackrel{\text{\sout{NEC}}}{\implies} & \text{TWH} \\
\Downarrow & & & & \\
\end{array}
$$
$$
\left(\sout{\text{NEC}} \,\&\, \exists\,\text{CTCs}\right) \,||\, \sout{\text{SC}}. \tag{3.6.1}
$$

Here, C denotes "quantum cloning", "QE" quantum entanglement, "SLS" superluminal signaling, "T" topology, "WH" wormholes, "TWH" traversable wormholes, and "SC" strong causality. The single-lined arrows in Eq. (3.6.1) indicate duality of specific statements under ER=EPR, double-lined arrows indicate logical implication, and strikethroughs indicate violation.

It is striking that on both the general relativistic and quantum mechanical sides of the duality, violation of the no-go theorem leads to problems for causality. The unexpected connection between cloning and topology change offers support for the ER=EPR correspondence, which provides a natural explanation for their relation.

A promising avenue for future research is the investigation of whether other no-go theorems in quantum mechanics and gravity neatly correspond under ER=EPR. The no-deleting theorem corresponds to the topology theorem in exactly the same way as the no-cloning theorem, while the no-communication theorem is equivalent to the assertion of nontraversability of wormholes. On the gravity side, violation of Hawking's area theorem, i.e., the generalized second law of thermodynamics, requires either breakdown of cosmic censorship or of the null energy condition [97], the latter allowing wormhole traversal [93]. In ER=EPR, this corresponds to violation of the no-communication theorem [16] and, in AdS/CFT, would correspond to violation of unitarity in the dual CFT state of Eq. (3.3.1) [96]. Whether all known gravitational or quantum mechanical no-go theorems map onto each other in this way is a fascinating open question. More generally, the connections among infrared constraints on ultraviolet physics, such as unitarity and causality [96, 98–100], will continue to play an important role in understanding quantum gravity.

Wormhole and Entanglement (Non-)Detection in the ER=EPR Correspondence

4.1   Introduction

Black holes are the paradigmatic example of a system where both field-theoretic and gravitational considerations are important. Black hole thermodynamics and the area theorem [57, 58] already provided a relationship between entanglement and geometry, while the classic black hole information paradox [65] and its potential resolution via complementarity [88, 101, 102] pointed at the subtlety of the needed quantum mechanical description. In the last few years, the firewall paradox [67] has heightened the tension between these two descriptions, prompting a number of proposals to modify the standard picture to a greater or lesser extent.

One set of proposals [104–107, 267] modifies quantum mechanics to allow for state-dependence of the black hole horizon, so that an infalling observer does not encounter a firewall even though the state of the black hole horizon can be written as a superposition of basis states that each have high energy excitations [108]. In order to avoid this problem, the presence or absence of a black hole firewall must become a nonlinear observable, contrary to standard quantum mechanics. In a recent paper, Marolf and Polchinski [109] have pointed out that this nonlinearity cannot be "hidden"; if it is strong enough to remove a firewall from a generic state, it must lead to violations of the Born Rule visible from outside the horizon.

In this paper, we consider a different idea inspired by the firewall paradox, the ER=EPR correspondence [16], which asserts the existence of an exact duality between Einstein-Podolsky-Rosen (EPR) pairs, i.e., entangled qubits, and Einstein-Rosen (ER) bridges [73–75], i.e., nontraversable wormholes. This duality is supposed to be contained within quantum gravity, which is in itself meant to be a *bona fide* quantum mechanical theory in the standard sense. The ER=EPR proposal is radical, but it is not obviously excluded by either theory or observation, and indeed has passed a number of nontrivial checks [69, 70, 110–112]; if true, it has the potential to relate previously unconnected statements about entanglement and general relativity in a manner reminiscent of the AdS/CFT correspondence [12–14]. In a previous paper [113], we pointed out that in ER=EPR the no-cloning theorem is dual to the general relativistic no-go theorem for topology change [77, 79]; violation on either side of the duality, given an ER bridge (two-sided black hole), would lead to causality violation and wormhole traversability.

In light of the result of Ref. [109], one might be worried that ER=EPR is in danger. It is well-known that entanglement is not an observable, in the sense we will make precise below; we cannot look at two spins and determine whether they

are in an arbitrary, unspecified entangled state with one another. Yet ER=EPR implies that the two spins are connected by a wormhole, so that the geometry of spacetime differs according to whether or not they are entangled. If this difference in geometry could be observed, entanglement would become a (necessarily nonlinear) observable as well and the laws of quantum mechanics would be violated, contrary to the assumption that the latter are obeyed by quantum gravity.

In this paper, we show that ER=EPR does not have this issue. Unlike the modifications to quantum mechanics considered by Ref. [109], wormhole geometry *can* be hidden. In particular, we show that in general relativity no measurement can detect whether the interior of a black hole has a wormhole geometry. More precisely, observers can check for the presence or absence of specific ER bridge configurations, but there is no projection operator (i.e., observable) onto the entire family of wormhole geometries, just as (and, in ER=EPR, for the same reason that) there is no projection operator onto the family of entangled states.

The remainder of this paper is organized as follows. We first review the basic quantum mechanical statement that entanglement is not an observable. Next we introduce the maximally extended AdS-Schwarzschild geometry in general relativity and, using AdS/CFT, on the CFT side. As a warmup, we first show that no single observer can detect the presence of a wormhole geometry. We then turn to more complicated multiple-observer setups and show, as desired, that they are unable to detect the presence of nontrivial topology in complete generality.

4.2 Entanglement Is Not an Observable

The proof that one cannot project onto a basis of entangled states [114] proceeds as follows. Assume the existence of a complete basis set of entangled states $|\psi_{E_i}\rangle$, distinct from the basis set of all states. A projection onto this basis could be written in the form

$$\hat{P}_E = \sum_i |\psi_{E_i}\rangle\langle\psi_{E_i}|. \tag{4.2.1}$$

Note, however, that the set of all entangled states has support over the entire Hilbert space, as the entangled states can be written as linear sums of unentangled states:

$$|\psi_{E_i}\rangle = \sum_{j \in B_i} |\psi_j\rangle \tag{4.2.2}$$

for some set B_i. Therefore, the projector onto the set of all entangled states does not project out any states in the Hilbert space. Said another way, the set of all entangled states is not a set that is closed under superposition, thus preventing a projection thereupon. Since no projector exits, entanglement is therefore not an observable.

4.3 Setup

We consider the maximally extended AdS-Schwarzschild geometry [115, 116], which, following Ref. [16], we will interpret as an Einstein-Rosen bridge connecting two black holes. The metric for the AdS-Schwarzschild black hole in D spacetime dimensions is [117, 118]

$$ds^2 = -f(r)dt^2 + \frac{dr^2}{f(r)} + r^2 d\Omega_{D-2}^2, \tag{4.3.1}$$

where $d\Omega_{D-2}^2$ is the surface element of the unit $(D-2)$-dimensional sphere and $f(r)$ is defined to be

$$f(r) = 1 - \frac{16\pi G_D M}{(D-2)\Omega_{D-2} r^{D-3}} + \frac{r^2}{L^2}, \tag{4.3.2}$$

writing G_D for Newton's constant in D dimensions, $\Omega_{D-2} = 2\pi^{(D-1)/2}/\Gamma[(D-1)/2]$ for the area of the unit $(D-2)$-sphere, and L for the AdS scale. The horizon r_{H} is located at the point where $f(r_{\mathrm{H}}) = 0$. The tortoise coordinate can be defined as $r^* = \int dr/f(r)$, the ingoing and outgoing Eddington-Finkelstein coordinates $v = t + r^*$ and $u = t - r^*$, with which we can define the lightcone Kruskal-Szekeres coordinates

$$
\begin{aligned}
\text{(I)} \quad & U = -e^{-f'(r_{\mathrm{H}})u/2} & V = e^{f'(r_{\mathrm{H}})v/2} \\
\text{(II)} \quad & U = e^{-f'(r_{\mathrm{H}})u/2} & V = e^{f'(r_{\mathrm{H}})v/2} \\
\text{(III)} \quad & U = e^{-f'(r_{\mathrm{H}})u/2} & V = -e^{f'(r_{\mathrm{H}})v/2} \\
\text{(IV)} \quad & U = -e^{-f'(r_{\mathrm{H}})u/2} & V = -e^{f'(r_{\mathrm{H}})v/2}.
\end{aligned}
\tag{4.3.3}
$$

Regions I through IV are depicted in Fig. 4.1 and define the maximally extended AdS-Schwarzschild black hole geometry. Defining $T = (U+V)/2$ and $X = (V-U)/2$, the horizon is located at $T = \pm X$, that is, at $UV = 0$, while the singularity is located at $T^2 - X^2 = 1$. The one-sided AdS black hole occupies Region I and half of Region
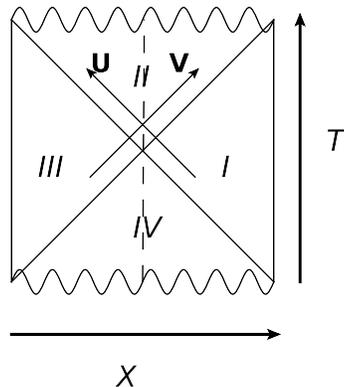
Figure 4.1. The maximally extended AdS-Schwarzschild geometry, with Kruskal-Szekeres coordinates T, X and lightcone coordinates U, V indicated. Of course, the singularity actually appears as a hyperbola in T, X. This diagram is a conformally-transformed sketch to indicate the general relationship among the coordinates; see Ref. [119] for more discussion. Regions I through IV are defined by Eq. (4.3.3).

II, i.e., $V > 0, X > 0$. In these coordinates, the metric becomes

$$
\begin{aligned}
\mathrm{d}s^2 &= -\frac{4|f(r)|e^{-f'(r_{\mathrm{H}})r^*}}{[f'(r_{\mathrm{H}})]^2}\mathrm{d}U\mathrm{d}V + r^2\mathrm{d}\Omega_{D-2}^2 \\
&= \frac{4|f(r)|e^{-f'(r_{\mathrm{H}})r^*}}{[f'(r_{\mathrm{H}})]^2}(-\mathrm{d}T^2 + \mathrm{d}X^2) + r^2\mathrm{d}\Omega_{D-2}^2,
\end{aligned}
\tag{4.3.4}
$$

where r is now defined implicitly in terms of U and V via

$$
UV = T^2 - X^2 = \pm e^{f'(r_{\mathrm{H}})r^*},
\tag{4.3.5}
$$

where the sign is $-$ for Regions I and III and $+$ for Regions II and IV.

We now turn to the CFT interpretation of the geometry. In the Maldacena and Susskind proposal of ER=EPR [16], it is pointed out that, in AdS/CFT, the state $|\psi(t)\rangle$ in the CFT corresponds at different times to different causal diamonds in the eternal, maximally extended AdS-Schwarzschild geometry. Different spatial slices through a given causal diamond that intersect the boundaries at fixed points are related to each other by the Wheeler-deWitt equation in the bulk. If one is outside a black hole in AdS, without knowing *a priori* which time slice one is on, then the different $|\psi(t)\rangle$ are simply a one-parameter family of states $|\psi_\alpha\rangle$, where α has replaced t and is now just the label for the state of the CFT at time $t = 0$; all of
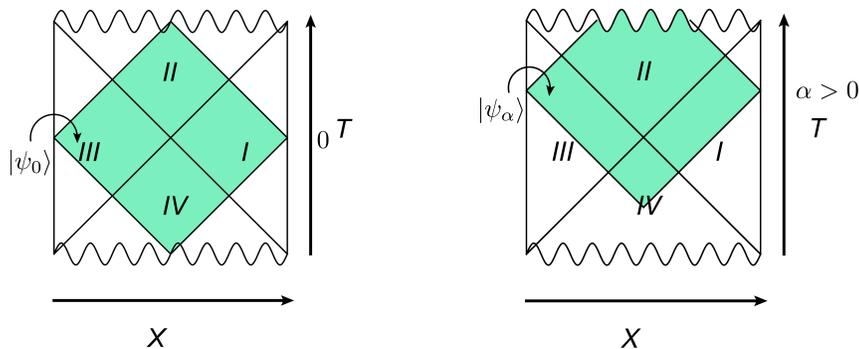
Figure 4.2. (left) The state $|\psi_0\rangle$, corresponding to a wormhole geometry where the ER bridge intersects the boundary at $T = 0$. (right) The family of states $|\psi_\alpha\rangle$, $\alpha > 0$, for which the ER bridge intersects the boundary at $T > 0$.

the $|\psi_\alpha\rangle$ describe pairs of black holes containing some sort of ER bridge. The various geometries are shown in Fig. 4.2.

Famously, the maximally extended AdS black hole can be described on the CFT side of the AdS/CFT correspondence by the thermofield double state of two noninteracting large-N CFTs on the boundary sphere. We take the interpretation [16] of the state as two entangled black holes that both evolve forward in time, that is,

$$|\psi(t)\rangle = \frac{1}{\sqrt{Z}} \sum_n e^{-\beta E_n/2} e^{-2iE_n t} |\bar{n}\rangle_L \otimes |n\rangle_R, \tag{4.3.6}$$

where $|n\rangle_L$ and $|n\rangle_R$ are the nth eigenstates of the left and right CFTs, respectively, with eigenvalue E_n, a bar denotes the CPT conjugate, and β is the inverse temperature. We note that the CFT time t in Eq. (5.4.1) is the $r \to \infty$ limit of the Schwarzschild time t that appears in Eq. (4.3.1). By considering the surface of constant Kruskal time T that intersects the $r = \infty$ boundary at Schwarzschild time t, we can instead parameterize the CFT state corresponding to the eternal AdS black hole as $|\psi(T)\rangle$. Equivalently, we can write as $|\psi_T\rangle$ the family of ER bridges indexed by T, which correspond at the fixed Kruskal time $T = 0$ to the CFT state $|\psi(T)\rangle$. The black hole described by $|\psi_{T_0}\rangle$ is given by the metric (4.3.4) with T replaced with $T - T_0$ in Eq. (4.3.5). The analogous states with two one-sided black holes on the boundary CFTs will be called $|\phi_T\rangle$, where

$$|\phi_t\rangle = \frac{1}{\sqrt{Z}} \left(\sum_m e^{-\beta E_m/2} e^{-iE_m t} |\bar{m}\rangle_L \right) \otimes \left(\sum_n e^{-\beta E_n/2} e^{-iE_n t} |n\rangle_R \right). \tag{4.3.7}$$

4.4 The Single-Observer Case

To gain intuition for the setup, in this section we restrict ourselves to measurements that a single (test particle) observer can perform in an otherwise empty (AdS-)Schwarzschild spacetime. Such observers are forbidden from receiving information from or coordinating with other observers; that is, we first investigate the aspects of the geometry that can be probed by a single causal geodesic. We will refer to this class of observers as *isolated* observers. The simplest way for an isolated observer to verify the existence of an ER bridge would be to pass through it, i.e., to traverse the wormhole. It turns out, however, that this process is disallowed both by classical general relativity and, via ER=EPR, by quantum mechanics.

In general relativity, the nontraversability of wormholes follows immediately from a more fundamental result, the topological censorship theorem [120], which is the statement that in a globally hyperbolic, asymptotically flat spacetime satisfying the null energy condition (NEC), any causal curve from past null infinity to future null infinity is diffeomorphic to an infinite causal curve in topologically trivial spacetime (such as Minkowski space). In other words, no causal observer's worldline can ever probe nontriviality of topology of spacetime.[1] Probing the nontrivial topology of an ER bridge simply means passing through the wormhole, which is therefore forbidden given the NEC. In a previous paper [113], we showed that violation of the NEC in ER=EPR necessarily leads to violation of the no-cloning theorem and the breakdown of unitary evolution. Traversable wormholes are therefore also forbidden by quantum mechanics given ER=EPR, as they would correspond to a breakdown of unitarity by allowing superluminal signaling.

The next simplest means of verifying the existence of an ER bridge would be to detect the nontrivial topology of the wormhole without traversing it. In the present context, we see that detecting the nontrivial topology is equivalent under ER=EPR to detecting the existence of entanglement—more precisely, to constructing a linear operator that detects if an unknown state is entangled with anything else. But it is well known that such an operator is forbidden by the linearity of quantum mechanics,

---

[1] Of course, nonisolated observers *can* determine topological characteristics of their spacetime, for example by seeing the same stars on opposite sides of the sky and thereby determining that spatial sections of their spacetime are toroidal. However, they must receive information from outside their worldline—in this case, photons emitted by distant stars that travel on topologically distinct geodesics—to do so. Furthermore, the topological censorship theorem guarantees that if the spacetime is asymptotically flat, satisfies the NEC, and allows Cauchy evolution, then any handles must collapse to a singularity before an observer can travel around them.

as Ref. [109] discusses. Briefly, this is because projection operators cannot project onto a subspace unless that subspace is closed under superposition. An attempt to project onto the set of all entangled states will therefore fail due to the set of all entangled states not being closed under superposition; such a projector will inevitably project onto the entire Hilbert space of all states. On the gravity side, this leads to a result stronger than the nontraversability of wormholes: not only does ER=EPR forbid an observer from traversing wormholes, it forbids an isolated observer from verifying their existence even once inside them.

This result can be straightforwardly verified in general relativity by examining the applicable metrics. Importantly, the metric given in Eqs. (4.3.4) and (4.3.5) for the maximally extended geometry has several isometries: it is invariant under the exchange $(U, V) \leftrightarrow (-U, -V)$ and also under the exchanges $(T, X) \leftrightarrow (T, -X)$ and $(T, X) \leftrightarrow (-T, X)$. That is, Regions I and II in Fig. 4.1 are the same as Regions III and IV, respectively, and moreover the entire metric is symmetric under spatial (X) or temporal (T) reversal. In particular, the regions present in both this geometry and the one-sided black hole geometry (Region I and half of Region II, i.e., $V > 0, X > 0$) are completely identical in the two cases. It is this property that implies that an observer on a geodesic entering a one-sided black hole cannot distinguish it from a two-sided black hole via any local measurement of curvature.

We have therefore shown that a single (isolated) observer cannot observe whether a given black hole hosts an ER bridge, even by jumping into it. We next consider observables that require multiple communicating observers to implement.

4.5   The Multiple-Observer Case

One can ask the question of whether two (or, for that matter, many) observers can detect the existence of entanglement or, equivalently, of nontrivial topology. The setup of the experiment is as follows. Consider a maximally extended, eternal AdS-Schwarzschild geometry, as depicted in Fig. 4.3. Allow two observers, Alice and Bob, to initially begin in the white hole portion of the geometry. (We will consider the case of more than two observers later in this section.) Now let the observers exit the
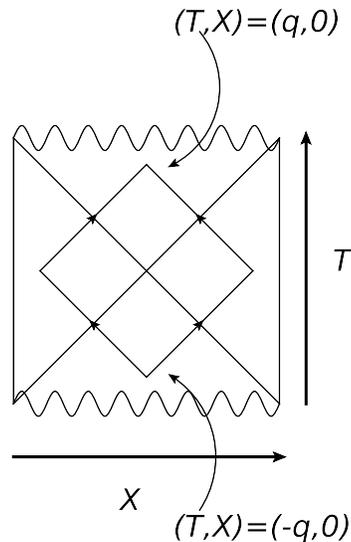
Figure 4.3. The procedure described in the text for detecting a wormhole. Alice and Bob emerge from the white hole portion of the AdS-Schwarzschild geometry, then meet again inside the black hole.

white hole[2] to the two different asymptotic regions not contained in the black hole. Next, they both jump into their respective black holes and compare notes. In such a way, they could potentially determine if there was entanglement before hitting the singularity.

The problem with this construction is that it doesn't definitively tell the observers if there was entanglement or not. Indeed, Alice and Bob could jump into the $|\psi_0\rangle$ ER bridge at sufficiently late time T that they are unable to communicate (since one or both of them will hit the singularity before being able to do so); equivalently, the geometry could be $|\psi_\alpha\rangle$ for α too large (instead of $|\psi_0\rangle$), as depicted in Fig. 4.4. The same argument that states that no linear operator permits the observers from detecting whether or not there is entanglement precludes this verification procedure from succeeding with probability 1. But how is it possible to reconcile the fact that Alice and Bob can sometimes verify the existence of an ER bridge with the

---

[2]We note that the white holes mentioned in our construction are for convenience only; it suffices for Alice and Bob to have communicated at some past time and simply to have moved out of causal contact. Indeed, it is possible for Alice and Bob to both exist in the same asymptotically AdS vacuum, as long as a wormhole exists connecting their locations. It is, however, necessary for them to enter the wormhole in order to attempt to detect information regarding the entanglement in this picture.
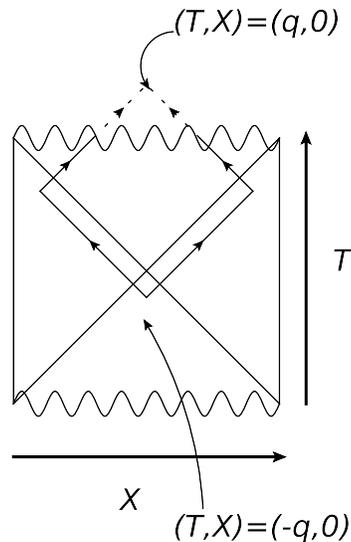
Figure 4.4. Unlike in Fig. 4.3 above, here the geometry is shifted to some $|\psi_\alpha\rangle$ for some sufficiently large $\alpha \neq 0$; Alice and Bob hit the singularity before they can meet and are therefore unable to verify the existence of a wormhole.

impossibility of projecting onto a generic family of states?

To be concrete, suppose that Alice and Bob are in some geometry in the set of all $|\psi_\alpha\rangle$ (for α unknown) and travel on outgoing nearly null trajectories beginning in the white hole at $(T, X) = (-q, 0)$, with Alice (Bob) entering Region III (respectively, I) at $(T, X) \simeq (-q/\sqrt{2}, \mp q/\sqrt{2})$, turning around at $(T, X) \simeq (0, \mp q)$, and entering Region II through their respective black holes at $(T, X) \simeq (q/\sqrt{2}, \mp q/\sqrt{2})$. Now, if they do not hit a singularity, their geodesics will cross again at $(T, X) \simeq (q, 0)$. That is, their geodesics will cross if they are in the state $|\psi_\alpha\rangle$, i.e., the state in which T is shifted by α, for $\alpha < 1 - q$. If $\alpha > 1 - q$ for a $|\psi_\alpha\rangle$ state or if they had instead been in any one of the $|\phi_\alpha\rangle$ states, they would hit the singularity without their paths ever crossing. (Recall that for $X = 0$, the singularity is located at $T = \pm 1$ for the $|\psi_0\rangle$ geometry.) Hence, Alice and Bob are able to verify if they are in the set $S_q = \{|\psi_\alpha\rangle | \alpha < 1 - q\}$.

However, this thought experiment does not require the existence of a projection operator onto the entire family S_q. Instead, after their geodesics cross, Alice and Bob can actually determine in which of the $|\psi_\alpha\rangle$ they are. All null geodesics from the horizon to the singularity are isomorphic and experience the same pattern of values of the curvature tensor on the way in. That is, a family of null geodesics with,

e.g., constant $V = V_0$ can be labeled by the time t at which they cross a surface at fixed proper distance from the horizon in Region I, which is the only difference among the geodesics; since the metric (4.3.1) is independent of t, all of these geodesics experience the same inward journey. Hence, before meeting Bob inside Region II, there is no distinguishing event by which Alice can measure α. However, the value of the Riemann tensor *at the moment Alice's and Bob's geodesics cross* is unique for each $|\psi_\alpha\rangle$.

In particular, at the moment their geodesics cross, Alice and Bob can measure the tidal forces acting in their local Lorentz frames by computing some component of the Weyl tensor. At $(T, X) = (q, 0)$, Eq. (4.3.5) implies that, in Region II, r^* and hence r is a monotonic function of $(q - \alpha)$. (Recall that for $|\psi_\alpha\rangle$ all equations describing the metric are shifted by $T \to T - \alpha$.) Let us define a local Lorentz frame in coordinates $(\hat{t}, \hat{r}, \hat{\theta})$, where $\hat{\theta}$ is the orthonormal coordinate in the $D - 2$ angular directions. When their paths cross, Alice and Bob can measure the $\hat{r}\hat{\theta}\hat{r}\hat{\theta}$ component of the Weyl tensor, which is

$$W_{\hat{r}\hat{\theta}\hat{r}\hat{\theta}} = -\frac{1}{L^2} - \left(\frac{D-3}{D-2}\right) \frac{8\pi G_D M}{\Omega_{D-2} r^{D-1}}. \tag{4.5.1}$$

Note that this quantity monotonically increases as r [and so in $(q - \alpha)$]. This implies that Alice and Bob can determine α by measuring tidal forces at the moment when their geodesics cross; there is a bijection between α and the size of the tidal force. This measurement thus acts as a projection operator $P_\alpha = |\psi_\alpha\rangle\langle\psi_\alpha|$. This is analogous to the possibility of being able to detect if two qubits are in some particular entangled state, rather than absolutely any entangled state whatsoever.

The key point here is that if the observers hit the singularity before exchanging a signal, i.e., if the wavefunction is one of the $|\psi_\alpha\rangle$ for which $\alpha > 1 - q$, then Alice and Bob are unable to confirm the existence of the ER bridge. If $\alpha < 1 - q$, the experiment Alice and Bob perform actually determines α. This procedure therefore fails to determine if the region behind a horizon contains a generic wormhole: it can sometimes reveal its existence, but not rule out its presence. It therefore does not implement a projector onto the set of all wormhole states. Thus, no contradiction with linearity of quantum mechanics arises in ER=EPR from the ability of Alice and Bob to jointly explore the wormhole geometry.

A priori, one could wonder whether even more general configurations of more than two observers could make the existence of wormhole topology into an observable. Note that it is not consistent to consider a setup in which there is an infinite set of

observers (or signals) entering a horizon at earlier and earlier times, as this would violate the necessary assumption of weak backreaction and hence invalidate the AdS-Schwarzschild spacetime ansatz. Hence, in a given slicing of spacetime, there must be an initial observer to enter the horizon. A prototypical setup for the thought experiment with more than two observers can therefore be rephrased as follows. After meeting and arranging the experiment, Bob and Alice go their separate ways. Bob jumps into his horizon, crossing it at spacetime point $p = (T, X) \simeq (q, q)/\sqrt{2}$ as before. This time, however, Alice remains outside her horizon and instead sends into her black hole multiple light pulses at regular intervals, with the first light pulse she emits (after leaving Bob) entering her wormhole mouth at $p' = (T, X) \simeq (q', -q')/\sqrt{2}$. The multiple light pulses are equivalent to having multiple observers enter the black hole at different times. However, one can choose a slicing of spacetime in which p and p' are on the same spacelike sheet; that is, one can simply apply a boost to equate the spatial components of p and p'. Since a boost can be independently applied to each asymptotically AdS spacetime, it follows that the case in which Bob is also replaced by multiple observers can be similarly simplified. As a result, the multiple observer setup reduces to the two observer setup, which we showed previously cannot definitively answer the question of whether there is a wormhole geometry.

Thus, even with multiple observers, the measurement of whether or not there is an ER bridge in general is not a valid observable, any more than the question of whether two qubits are arbitrarily entangled is a quantum mechanical observable.

4.6 Conclusions

The ER=EPR proposal is a compelling but surprising idea about quantum gravity, identifying features of ordinary quantum mechanics with geometrical and topological features of spacetime. As an extraordinary claim, it is necessary that it be subjected to rigorous theoretical tests to ascertain whether it suffers from any inconsistencies. One such potential issue, which we have addressed in this paper, is whether ER=EPR implies a serious modification of quantum mechanics, namely, the introduction of state dependence. The argument that ER=EPR implies state dependence rests on the observation that the correspondence identifies entanglement with wormholes. Famously, entanglement is not a quantum mechanical observable, so this leads to the question of whether the observation of a wormhole contradicts, under ER=EPR, linearity of quantum mechanics.

In this paper, we have argued that ER=EPR does not contradict this principle of quantum mechanics precisely because the general question of the existence or nonexistence of a wormhole is also not an observable. We showed that neither a single observer nor a group of observers is able to definitively establish whether a pair of event horizons is linked by an ER bridge. A single observer can never detect the (nontraversable) wormhole's existence, which mirrors the fact that, given a single qubit, one cannot tell if it is entangled by anything else. On the other hand, by exploring the spacetime, two or more observers working in concert can decide if they are in a particular ER bridge geometry, but cannot project onto the entire family. Under ER=EPR, this statement mirrors the fact that one can project two qubits onto a particular entangled state but not onto the family of all possible entangled states.

Many options are available for future investigation. The ER=EPR correspondence has been subjected to some tests [69, 70, 110–113], but the challenge of seeing the duality between wormholes and any arbitrary form of quantum entanglement remains, as does the very definition of what is meant by a "wormhole" in ER=EPR for theories without a weakly-coupled holographic gravity dual. Other open issues include the investigation of whether firewalls are truly nongeneric in ER=EPR [73] and whether the correspondence can be concretely realized outside of asymptotically AdS spacetime. The answers to these questions and others will likely provide important insight in future investigations in the connections between entanglement and spacetime geometry.

Acknowledgments

Entanglement Conservation, ER=EPR, and a New Classical Area Theorem for Wormholes

5.1 Introduction

All of the states of a quantum mechanical theory are on the same footing when considered as vectors in a Hilbert space: any state can be transformed into any other state by the application of a unitary operator. When the Hilbert space can be decomposed into subsystems, however, there is a natural way to categorize them: by the entanglement entropy of the reduced density matrix of a subsystem constructed from the states. Entanglement between two subsystems is responsible for the "spooky action at a distance" often considered a characteristic feature of quantum mechanics: measuring some property of a subsystem determines the outcome of measuring the same property on another entangled subsystem, even a causally disconnected one.

It is well known that this seeming nonlocality does not lead to violations of causality. It cannot be used to send faster-than-light messages [84] and in fact it is impossible for any measurement to determine whether the state is entangled (see, e.g., Ref. [114]). Similarly, it is impossible to alter the entanglement between a system and its environment (that is, to change the entanglement entropy of the reduced density matrix of the system) by acting purely on the degrees of freedom in the system or by adding more unentangled degrees of freedom. A number of well-established properties, such as monogamy [121] and strong subadditivity [122], constrain the entanglement entropy of subsystems created from arbitrary factorizations of the Hilbert space.

Although entanglement entropy is a fundamental quantity, it is typically very difficult to compute in field theories, where working directly with the reduced density matrix can be computationally intractable, although important progress has been made in certain conformal field theories [123, 124] and more generally along lightsheets for interacting quantum field theories [125]. The AdS/CFT correspondence [12–14], however, allows us to transform many field-theoretic questions to a gravitational footing. In particular, the Ryu-Takayanagi formula [15] equates the entanglement entropy of a region for a state in a conformal field theory living on the boundary of an asymptotically AdS spacetime to the area of a minimal surface with the same boundary as that region in the spacetime corresponding to that CFT state. Using this identification of entropy with area, a number of "holographic entanglement inequalities" have been proven [126, 127], some reproducing and some stronger than the purely quantum mechanical entanglement inequalities.

Motivated in part by AdS/CFT, as well as a number of older ideas in black hole thermodynamics [57, 58] and holography [59, 60, 128], Maldacena and Susskind

have recently conjectured [16] an ER=EPR correspondence, an exact duality between entangled states (Einstein-Podolsky-Rosen [71] pairs) and so-called "quantum wormholes", which reduce in the classical general relativistic limit to two-sided black holes (Einstein-Rosen [73] bridges, i.e., wormholes). In a series of recent papers, we have considered the implications of this correspondence in the purely classical regime. In this limit, if the ER=EPR duality holds true, certain statements in quantum mechanics about entangled states should match directly with statements in general relativity about black holes and wormholes [129], with the same assumptions required on both sides. We indeed previously found two beautiful and nontrivial detailed correspondences: the no-cloning theorem in quantum mechanics corresponds to the no-go theorem for topology change in general relativity [130] and the unobservability of entanglement corresponds to the undetectability of the presence or absence of a wormhole [131].

In this paper, we extend this correspondence to a direct equality between the entanglement entropy and a certain invariant area, which we define, of a geometry containing classical black holes and wormholes. We follow a long tradition of clarifying general relativistic dynamics using area theorems [132–136], which hold that various areas of interest satisfy certain properties under time evolution. Our strategy is to show that the area in question remains unchanged under dynamics constituting the gravitational analogue of applying tensor product operators to an individual system and its complement. We show that, just as entanglement entropy cannot be changed by acting on the subsystem and its complement separately, this area is not altered by merging pairs of black holes or wormholes or by adding classical (unentangled) matter. The area we consider is chosen to be that of a maximin surface [137, 138] for a collection of wormhole horizons, a time-dependent generalization of the Ryu-Takayanagi minimal area, which again establishes that the entanglement entropy is also conserved under these operations. At least for asymptotically AdS spacetimes, our result constitutes an explicit characterization of the ER=EPR correspondence in the classical limit. Moreover, our theorem is additionally interesting from the gravitational perspective alone, as it constitutes a new area law within general relativity.

This paper is structured as follows. In Section 5.2, we review the simple quantum mechanical fact that entanglement is conserved under local operations. In Section 5.3, we define the maximin surface and review its properties. In Section 5.4, we prove our desired general relativistic theorem. Finally, we discuss the implications of our result and conclude in Section 5.5.

5.2 Conservation of Entanglement

Consider a Hilbert space $\mathcal{H}$ that can be written as a tensor product of two factors $\mathcal{H}_\mathrm{L}$ and $\mathcal{H}_\mathrm{R}$ to which we will refer as "right" and "left", though they need not have any spatial interpretation. For a state $|\psi\rangle \in \mathcal{H}$, let us define the reduced density matrix associated with $\mathcal{H}_\mathrm{L}$ as $\rho_\mathrm{L} = \mathrm{Tr}_{\mathcal{H}_\mathrm{R}} |\psi\rangle\langle\psi|$ and use this to define the entanglement entropy between the right and left sides of the Hilbert space:

$$S(L) = S(R) = -\mathrm{Tr}_{\mathcal{H}_\mathrm{L}} \rho_\mathrm{L} \log \rho_\mathrm{L}. \tag{5.2.1}$$

It is straightforward to see that adding more unentangled degrees of freedom to $\mathcal{H}_\mathrm{L}$ will not affect the entanglement entropy, as by construction this does not introduce new correlations between $\mathcal{H}_\mathrm{L}$ and $\mathcal{H}_\mathrm{R}$. This is particularly clear to see by using the equivalence of $S(L)$ and $S(R)$ for pure states, as adding in further unentangled degrees of freedom will maintain the purity of the joint system.

Now let us consider the effect on $S(L)$ of applying a unitary $U = U_\mathrm{L} \otimes U_\mathrm{R}$ to $|\psi\rangle$. As $\mathrm{Tr}_{\mathcal{H}_\mathrm{R}} U = U_\mathrm{L}$, we can consider only the action of U_L on ρ_L, as U_R acts trivially in $\mathcal{H}_\mathrm{L}$. This transforms $S(L)$ into

$$S(L) = -\mathrm{Tr}_{\mathcal{H}_\mathrm{L}} U_\mathrm{L} \rho_\mathrm{L} U_\mathrm{L}^\dagger \log \left(U_\mathrm{L} \rho_\mathrm{L} U_\mathrm{L}^\dagger \right). \tag{5.2.2}$$

One can at this point expand the logarithm by power series, with individual terms of the form

$$S_n(L) = -\mathrm{Tr}_{\mathcal{H}_\mathrm{L}} c_n U_\mathrm{L} \rho_\mathrm{L} U_\mathrm{L}^\dagger \left(\mathbb{1} - U_\mathrm{L} \rho_\mathrm{L} U_\mathrm{L}^\dagger \right)^n \tag{5.2.3}$$

for some real c_n. For each term in the expansion of the product, all but the first U_L and the last $U_\mathrm{L}^\dagger$ will cancel as $U_\mathrm{L}^\dagger U_\mathrm{L} = \mathbb{1}$. Finally, by cyclicity of the trace, the remaining U_L and $U_\mathrm{L}^\dagger$ will also cancel, leaving $S_n(L)$ invariant. Thus, $S(L)$ remains invariant under unitary transformations of the form $U = U_\mathrm{L} \otimes U_\mathrm{R}$. This is the statement of conservation of entanglement.

5.3 The Maximin Surface

A holographic characterization of the entanglement entropy begins with its calculation on a constant-time slice, where the Ryu-Takayanagi (RT) formula [15] holds:

$$S(H) = \frac{A_H}{4G\hbar}. \tag{5.3.1}$$

This relates the area A_H of the minimal surface subtending a region H to the entanglement entropy of that region with its complement. When the region is a complete boundary, this reduces to the minimal surface homologous to the region. For example, in a hypothetical static wormhole geometry, the entanglement entropy between the two ends would be given by the minimal cross-sectional area of the wormhole.

This method of computing entanglement entropy on a constant-time slice for static geometries was generalized by the Hubeny-Rangamani-Takayanagi (HRT) proposal [137]. The key insight here was that in general there do not exist surfaces that have minimal area in time, as small perturbations can decrease the area. The new proposal was that the area now scales as the smallest extremal area surface, as opposed to the minimal area. The homology condition mentioned previously remains in this prescription.

The maximin proposal [138] gives an explicit algorithm for the implementation of the HRT prescription. In the following definitions, we will closely follow the conventions used by Wall [138]. We define $C[H, \Gamma]$ to be the codimension-two surface of minimal area homologous to H anchored to ∂H that lies on any complete achronal (i.e., spacelike or null) slice Γ. Note that $C[H, \Gamma]$ can refer to any minimal area surface that exists on Γ. Next, the maximin surface $C[H]$ is defined as any of the $C[H, \Gamma]$ with the largest area when optimized over all achronal surfaces Γ. When multiple such candidate maximin surfaces exist, we refine the definition of $C[H]$ to mean any such surface that is a local maximum as a functional over achronal surfaces Γ. In the HRT proposal, the entanglement of H with its complement in the boundary is given by $S(H) = \text{area}[C[H]]/4G\hbar$.

As an example, for a wormhole geometry in which we are computing the entanglement entropy between the two horizons of the ER bridge, ∂H is trivial and the homology condition means that $C[H, \Gamma]$ is the surface of minimal cross-sectional area on an achronal surface Γ in the interior causal diamond of the horizons. Then the maximin surface $C[H]$ is a $C[H, \Gamma]$ with Γ chosen such that the area is maximized.

Such surfaces can be shown to exist for large classes of spacetimes and in particular $C[H]$ can be proven to be equal to the extremal HRT surface for spacetimes obeying the null curvature condition, which is given by

$$R_{\mu\nu}k^\mu k^\nu \geq 0, \tag{5.3.2}$$

where k^μ is any null vector and $R_{\mu\nu}$ is the Ricci tensor.[1] As HRT is a covariant method of calculating entanglement entropy, the maximin construction is therefore manifestly covariant as well.

Maximin surfaces in general have some further nice properties, proven in Ref. [138]: they have smaller area than the causal surface (the edge of the causal domain of dependence associated with bulk causality), they move monotonically outward as the boundary region increases in size, they obey strong subadditivity, and they also obey monogamy of mutual information, but not necessarily other inequalities that hold for constant-time slices [126, 127, 138]:

$$S(AB) + S(BC) \geq S(B) + S(ABC),$$
$$S(AB) + S(BC) + S(AC) \geq S(A) + S(B) + S(C) + S(ABC) \tag{5.3.3}$$

for disjoint regions A, B, and C. The above statements are all proven in detail for maximin surfaces in Ref. [138].

5.4 A Multi-Wormhole Area Theorem

We are now ready to find the gravitational statement dual to entanglement conservation. Let us take as our spacetime M the most general possible setup to consider in the context of the ER=EPR correspondence: an arbitrary, dynamical collection of wormholes and black holes in asymptotically AdS spacetime. We work in D spacetime dimensions. Throughout, we will assume that M obeys the null curvature condition (5.3.2). The degrees of freedom associated with the Hilbert space $\mathcal{H} = \otimes_i \mathcal{H}_i$ can be considered to be localized on the union of the stretched horizons, with each horizon comprising one of the $\mathcal{H}_i$ factors. We choose our spacetime setup such that the wormholes are past-initialized, by which we mean that for $t \leq 0$ the wormholes are far apart and the spacetime around the wormholes is in vacuum, with negligible back-reaction. Suppose we arbitrarily divide this system into two subsystems by labeling each horizon as "left" or "right". The left and right Hilbert spaces factorize as $\mathcal{H}_{\mathrm{L}} = \otimes_i \mathcal{H}_{\mathrm{L},i}$ and $\mathcal{H}_{\mathrm{R}} = \otimes_i \mathcal{H}_{\mathrm{R},i}$, where $H_{\mathrm{L(R)},i}$ contains the degrees of freedom associated with horizon i in the left (right) set. Now, some of the black holes in the left subset may be entangled with each other and so be described by ER bridges among the left set. A similar statement applies to the right set. Importantly, there may be

---

[1] For spacetimes satisfying the Einstein equation $R_{\mu\nu} - R\,g_{\mu\nu}/2 = 8\pi G\,T_{\mu\nu}$ for energy-momentum tensor $T_{\mu\nu}$, the null curvature condition is equivalent to the null energy condition $T_{\mu\nu}k^\mu k^\nu \geq 0$.

horizons in the left set entangled with horizons in the right set, describing ER bridges across the left/right boundary. For the sake of tractability, we consider horizons that are only pairwise entangled and that begin in equal-mass pairs in the asymptotically AdS spacetime; this stipulation can be made without loss of generality provided we consider black holes smaller than the AdS length and do not consider changes to the asymptotic structure of the spacetime (see, *e.g.*, Ref. [139]). (To treat wormholes with mouths of unequal masses, we could start in an equal-mass configuration and add matter into one of the mouths.) We thus take any two horizons i and j that are entangled to be in the thermofield double state at $t = 0$,

$$\Pi_i \Pi_j |\psi\rangle(t = 0) = |\psi_{i,j}\rangle(t = 0) = \frac{1}{\sqrt{Z}} \sum_n e^{-\beta E_n/2} |\bar{n}\rangle_i \otimes |n\rangle_j, \qquad (5.4.1)$$

where Π_i is a projector onto the degrees of freedom associated with $\mathcal{H}_i$, $1/\beta$ is the temperature, and $|n\rangle_i$ is the nth eigenstate of the CFT corresponding to the degrees of freedom in $\mathcal{H}_i$ with eigenvalue E_n.

Let us define a time slicing of the spacetime M into spacelike codimension-one surfaces Σ_t parameterized by a real number t that smoothly approaches the standard AdS time coordinate in the limit of spacelike infinity, where the metric is asymptotically AdS. The Σ_t are chosen to pass through the wormholes without coordinate singularities along the horizon (cf. Kruskal coordinates); see Fig. 5.1 for an example geometry. For the wormholes spanning the left and right subsets, we write as L_i and R_i the null codimension-one surfaces that form the outermost left and right apparent horizons, respectively, and define $L = \cup_i L_i$ and $R = \cup_i R_i$. Note that, since new apparent horizons can form outside of the initial apparent horizons, L_i and R_i are each not necessarily connected, but are the piecewise-connected union of the outermost connected components of the apparent horizons. On a given spacelike slice, an apparent horizon is a boundary between regions in which the outgoing orthogonal null congruences are diverging (untrapped) or converging (trapped) [133]. Of course, the indexing i may become redundant if horizons merge among the L_i or R_i. Let us define the restriction of the outermost apparent horizons to the constant-time slice Σ_t as the spacelike codimension-two surfaces $L_{t,i} = L_i \cap \Sigma_t$ and $R_{t,i} = R_i \cap \Sigma_t$ and similarly $L_t = L \cap \Sigma_t$ and $R_t = R \cap \Sigma_t$. Without loss of generality, we will use the initial spatial separation of the wormholes along with diffeomorphism invariance to choose the Σ_t and the parameterization of t such that Σ_0 intersects the codimension-two bifurcation surfaces $B_i \equiv L_{0,i} = R_{0,i}$ at which all the wormholes have zero length. The past-initialization condition then means that the wormholes are far apart in the
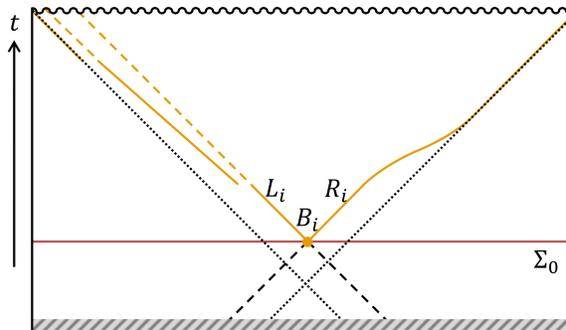
Figure 5.1. Penrose diagram, for an example spacetime M, of a slice through a particular wormhole i joining a left and right horizon. (Showing the full geometry would require a multi-sheeted Penrose diagram to accommodate the multiple wormholes.) The spacelike codimension-one surface Σ_0 is shown in burgundy. The initial bifurcation codimension-two surface B_i is illustrated by the orange dot. Apparent horizons are denoted by the orange lines, with the outermost apparent horizons L_i and R_i being the solid lines. For $t \leq 0$, the setup is past-initialized and the metric is given to good approximation by the eternal black hole in AdS, where the past event horizon of the white hole is indicated by the dashed black lines. The dotted black lines denote the future event horizon of M. As the spacetime at negative t is known, we do not show the entire Penrose diagram in this region, as indicated by the diagonal gray lines.

white hole portion of the spacetime, which corresponds to $t \leq 0$. Throughout, we will assume that $M \cup \partial M$ is globally hyperbolic; equivalently [140], we will assume that the closure of Σ_0 is a Cauchy surface for $M \cup \partial M$.

Now, for each $t > 0$, let us define a D-dimensional region of spacetime W_t as the union over all achronal surfaces with boundary $L_t \cup R_t$; that is, W_t is the causal diamond associated with $L_t \cup R_t$. A single wormhole has topology $S^{D-2} \otimes \mathbb{R}$ when restricted to Σ_t. The initial spacetime W_0 is special: it is a codimension-two surface that is just the union over all the B_i, with topology $(S^{D-2})^{\otimes N}$, where N is the number of wormholes connecting the left and right subsets.
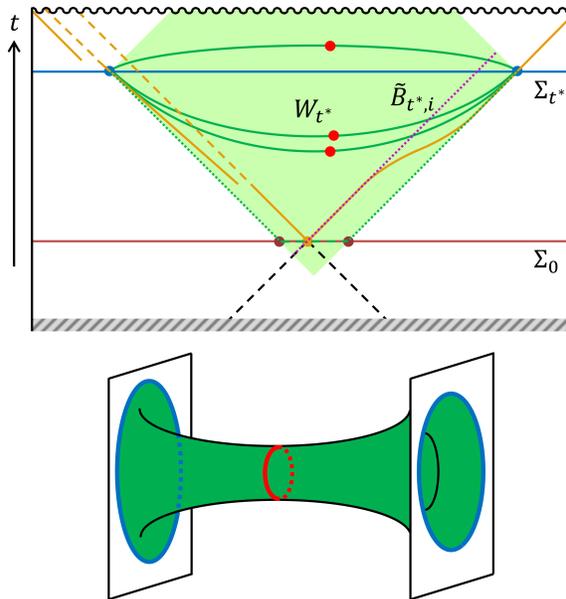
Figure 5.2. Penrose diagram (top), for the example geometry of Fig. 5.1, of the segment of the region W_{t^*} (green shading), for some t^*, that passes through a particular wormhole i joining a left and right horizon. The apparent horizons (orange lines, with solid lines for the outermost apparent horizons L_i and R_i), bifurcation surface B_i (orange dot), spacelike codimension-one surface Σ_0 (burgundy line), and past event horizons for the white hole (dashed black lines) are illustrated as in Fig. 5.1. The spacelike codimension-one surface Σ_{t^*} is shown as a blue line. The purple dotted line denotes the truncated null surface $\tilde{B}_{t^*,i}$ formed from the rightward outgoing orthogonal null congruence $\tilde{B}_i$ originating on B_i, used in Proposition 1. The codimension-two boundaries of W_{t^*} along wormhole i, $L_{t^*,i}$ and $R_{t^*,i}$, are indicated by the blue dots. The achronal codimension-one surfaces $\Gamma_{t^*}(\alpha)$ foliating W_{t^*} are indicated within wormhole i by the green lines; the codimension-two surfaces $C_{t^*}(\alpha)$ of minimal area for some slices $\Gamma_{t^*}(\alpha)$ are indicated within wormhole i by red dots. The particular surface $\Gamma_{t^*}(0)$, constructed in Eq. (5.4.8), is shown (for the portion restricted to wormhole i) by the dashed and dotted green lines, corresponding to $\Sigma_0 \cap W_{t^*}$ (the horizontal section) and $M_+ \cap \dot{J}^-[\Sigma_{t^*} \backslash W_{t^*}] = \tilde{L} \cup \tilde{R}$ (the diagonal sections), respectively. The burgundy dots denote the pieces of $\tilde{L}_0$ and $\tilde{R}_0$ in the vicinity of wormhole i. The embedding diagram (bottom) shows a particular slice $\Gamma_{t^*}(\alpha)$ through W_{t^*} for some α, where, as in the Penrose diagram, the codimension-two boundaries $L_{t^*,i}$ and $R_{t^*,i}$ are shown in blue and the surface $C_{t^*}(\alpha)$ of minimal cross-sectional area, restricted to wormhole i, is shown in red.

For a given W_t, let us define a slicing of W_t, parameterized by α, with achronal codimension-one surfaces $\Gamma_t(\alpha)$, where the boundary of $\Gamma_t(\alpha)$ is anchored at $L_t \cup R_t$ for all α and where α increases monotonically as we move from the past to the future

boundary of W_t. Now, we can imagine slicing $\Gamma_t(\alpha)$ into codimension-two surfaces and write as $C_t(\alpha)$ the surface with minimal area [i.e., the minimal cross-sectional area of $\Gamma_t(\alpha)$]; see Fig. 5.2. We can now define the maximin surface C_t for W_t as a surface for which the area of $C_t(\alpha)$ attains its maximum under our achronal slicing $\Gamma_t(\alpha)$, maximized over all possible such slicings. That is, C_t is a codimension-two surface with the maximum area, among the set of the surfaces of minimal cross-sectional area, for all achronal slices through W_t.

The main result that we will prove is that the area of the maximin surface C_t is actually independent of t, equaling just the sum of the areas of the initial bifurcation surfaces B_i.[2] In most cases, the maximin surface C_t will actually be the union of the initial bifurcation surfaces B_i, independent of t. In other words, the maximin area is invariant among all of the different causal diamonds W_t. Interpreting the area of the maximin surface as an entropy, this is the gravitational analogue of entanglement conservation. We will first prove a few intermediate results.

Proposition 1. The area of the maximin surface C_t is upper bounded by the sum of the areas of the initial bifurcation surfaces B_i.

Proof. Consider the rightward outgoing orthogonal null congruence $\tilde{B}_i$, a null surface of codimension one starting on B_i and satisfying the geodesic equation. Choosing some particular t^* arbitrarily, we truncate the null geodesics generating $\tilde{B}_i$ whenever a caustic is reached or when they intersect either the future singularity or the future null boundary of W_{t^*}; we further extend the null geodesics into the past until they intersect the past null boundary of W_{t^*}. We will hereafter write the truncated null surface as $\tilde{B}_{t^*,i}$. Let λ be an affine parameter for $\tilde{B}_{t^*,i}$ that increases toward the future and vanishes on B_i; let us write $\tilde{B}_{t^*,i}(\lambda)$ for the spatial codimension-two surface at fixed λ. The rotation $\hat{\omega}_{\mu\nu}$ in a space orthogonal to the tangent vector $k^\mu = (\mathrm{d}/\mathrm{d}\lambda)^\mu$ satisfies [142]

$$\frac{\mathrm{D}\hat{\omega}_{\mu\nu}}{\mathrm{d}\lambda} = -\theta\hat{\omega}_{\mu\nu}, \tag{5.4.2}$$

where $\theta = \nabla_\mu k^\mu$ is the expansion. Since θ vanishes on B_i, $\hat{\omega}_{\mu\nu}$ vanishes identically

---

[2]In Ref. [141] it was shown for the special cases of the Schwarzschild-AdS and the single, symmetric, Vaidya-Schwarzschild-AdS geometries that the initial bifurcation surface is the extremal surface in the HRT prescription. Our theorem in this paper generalizes this result to an arbitrary, dynamical, multi-wormhole geometry in asymptotically AdS spacetime that is past-initialized and that obeys the null curvature condition.

on $\tilde{B}_{t^*,i}$. The Raychaudhuri equation is therefore

$$\frac{\mathrm{d}\theta}{\mathrm{d}\lambda} = -\frac{1}{D-2}\theta^2 - \hat{\sigma}_{\mu\nu}\hat{\sigma}^{\mu\nu} - R_{\mu\nu}k^\mu k^\nu, \tag{5.4.3}$$

where $\hat{\sigma}_{\mu\nu}$ is the shear and $R_{\mu\nu}$ is the Ricci tensor. We note that if the null curvature condition (5.3.2) is satisfied, then θ is nonincreasing, as $\hat{\sigma}_{\mu\nu}\hat{\sigma}^{\mu\nu}$ is always nonnegative. Since the apparent horizon consists of marginally outer trapped surfaces (i.e., surfaces for which the outgoing orthogonal null geodesics have $\theta = 0$), it must be either null or spacelike, so any orthogonal null congruence starting on the apparent horizon remains either on or inside the apparent horizon in the future [133]. In particular, $\tilde{B}_{t^*,i} \subset W_{t^*}$.

Now, we can also write θ as $\mathrm{d}\log \delta A/\mathrm{d}\lambda$, where δA is an infinitesimal cross-sectional area element of $\tilde{B}_{t^*,i}(\lambda)$. That is, $\mathrm{area}[\tilde{B}_{t^*,i}(\lambda)]$ has negative second derivative in λ. Since θ vanishes on the bifurcation surface $B_i = \tilde{B}_{t^*,i}(0)$, we have that $\mathrm{area}[\tilde{B}_{t^*,i}(\lambda)]$ is monotonically nonincreasing in λ. Moreover, since for all $\lambda < 0$ there exists $t < 0$ such that $\tilde{B}_{t^*,i}(\lambda) \subset \Sigma_t$, the past-initialization condition means that $\mathrm{area}[\tilde{B}_{t^*,i}(\lambda)] = \mathrm{area}[B_i]$ for all $\lambda < 0$. Hence, for all λ we have

$$\mathrm{area}[\tilde{B}_{t^*,i}(\lambda)] \leq \mathrm{area}[B_i]. \tag{5.4.4}$$

By the past-initialization condition, there are no caustics to the past of B_i. Further, by definition, the wormhole does not pinch off until the singularity is reached, so some subset of the generators of $\tilde{B}_i$ must extend all the way through W_{t^*} without encountering caustics. Writing $\Gamma_{t^*}(\alpha)$ as a foliation of W_{t^*} by achronal slices, we thus have that $\tilde{B}_{t^*,i}(\lambda) \cap \Gamma_{t^*}(\alpha)$ is never an empty set for all α, i.e., for all λ there exists α such that $\tilde{B}_{t^*,i}(\lambda) \subset \Gamma_{t^*}(\alpha)$. Moreover, we can reparameterize and identify the affine parameters for each i of the $\tilde{B}_{t^*,i}$ such that for each λ there exists α for which $\cup_i \tilde{B}_{t^*,i}(\lambda) \subset \Gamma_{t^*}(\alpha)$; for such α, $\cup_i \tilde{B}_{t^*,i}$ is a complete cross-section of $\Gamma_{t^*}(\alpha)$, possibly with redundancy due to merging horizons. We choose our slicing $\Gamma_{t^*}(\alpha)$ such that there exists some α^* for which $\Gamma_{t^*}(\alpha^*)$ contains the maximin surface C_{t^*} for W_{t^*}, so

$$C_{t^*} = C_{t^*}(\alpha^*) \text{ such that } \mathrm{area}[C_{t^*}(\alpha^*)] = \max_\alpha \mathrm{area}[C_{t^*}(\alpha)], \tag{5.4.5}$$

where $C_{t^*}(\alpha)$ is the codimension-two cross-section of $\Gamma_{t^*}(\alpha)$ with minimal area.

Since $\tilde{B}_{t^*,i}$ is only completely truncated at future and past boundaries of W_{t^*}, it follows that for every α there must exist λ such that $\Gamma_{t^*}(\alpha) \supset \tilde{B}_{t^*,i}(\lambda)$. By the

definition of $C_{t^*}(\alpha)$, we have (for such λ) that

$$\text{area}[C_{t^*}(\alpha)] \leq \sum_i \text{area}[\tilde{B}_{t^*,i}(\lambda)]. \tag{5.4.6}$$

Putting together Eqs. (5.4.4) and (5.4.6), taking the maximum over λ and α on both sides, applying Eq. (5.4.5), and using the fact that t^* was chosen arbitrarily, we have a t-independent upper bound on the area of the maximin surface C_t:

$$\text{area}[C_t] \leq \sum_i \text{area}[B_i]. \tag{5.4.7}$$

$\square$

Let us now construct a lower bound on the area of the maximin surface C_t. We can do this by examining an achronal codimension-one surface through W_t and computing its minimal cross-sectional area; judiciously choosing the achronal surface optimizes the bound. In particular, for some arbitrary t^*, consider $\Gamma_{t^*}(0)$ passing through $\cup_i B_i$, where we choose the slicing such that

$$\Gamma_{t^*}(0) = (\Sigma_0 \cap W_{t^*}) \cup \left(M_+ \cap \dot{J}^-[\Sigma_{t^*} \backslash W_{t^*}] \right), \tag{5.4.8}$$

where M_+ is the restriction of M to $t \geq 0$, $J^-[A]$ denotes the causal past of a set A, and the dot denotes its boundary. That is, $\Gamma_{t^*}(0)$ consists of the codimension-one null surfaces forming the $t \geq 0$ portion of the boundary of W_{t^*} towards the past, plus a codimension-one segment of Σ_0 containing $\cup_i B_i$; see Fig. 5.2. Let us label the left and right boundaries of $\Sigma_0 \cup W_{t^*}$ (equivalently, the left and right portions of the intersection of Σ_0 and $\dot{J}^-[\Sigma_{t^*} \backslash W_{t^*}]$) as $\tilde{L}_0$ and $\tilde{R}_0$, respectively.

We will show in two steps that the minimal cross-sectional area of $\Gamma_{t^*}(0)$ is just $\sum_i \text{area}[B_i]$. We will first consider the cross-sectional area of slices of $\Sigma_0 \cap W_{t^*}$ and then examine the changes in cross-sectional area along slices of $M_+ \cap \dot{J}^-[\Sigma_{t^*} \backslash W_{t^*}]$.

Proposition 2. The minimal cross-sectional area of $\Sigma_0 \cap W_{t^*}$ is $\sum_i \text{area}[B_i]$.

Proof. By the requirement that the wormholes be past-initialized, the metric on Σ_0 is, up to negligible back-reaction, just a number of copies of the metric on the $t = 0$ slice of the single maximally-extended AdS-Schwarzschild black hole; for this metric the $t_{\text{KS}} = 0$ and $t_{\text{S}} = 0$ slices are the same, where t_{KS} is the Kruskal-Szekeres time coordinate and t_{S} is the Schwarzschild time coordinate [131]. Taking the t-slicing

to correspond to the Kruskal-Szekeres coordinates in the vicinity of each wormhole, therefore, the metric on $\Sigma_0 \cap W_{t^*}$ is

$$\mathrm{ds}^2_{\Sigma_0 \cap W_t} = \frac{4|f(r)|e^{-f'(r_\mathrm{H})r^*}}{[f'(r_\mathrm{H})]^2}\mathrm{d}X^2 + r^2\mathrm{d}\Omega^2_{D-2} = \frac{\mathrm{d}r^2}{f(r)} + r^2\mathrm{d}\Omega^2_{D-2}, \qquad (5.4.9)$$

where on Σ_0, the Kruskal X coordinate describing distance away from the wormhole mouth at B_i is $X = \pm e^{f'(r_\mathrm{H})r^*/2}$, with the sign demarcating the left and right side of B_i and the tortoise coordinate being $r^* = \int \mathrm{d}r/f(r)$. The function $f(r)$ is

$$f(r) = 1 - \frac{16\pi G_D M}{(D-2)\Omega_{D-2}r^{D-3}} + \frac{r^2}{\ell^2}, \qquad (5.4.10)$$

where Ω_{D-2} is the area of the unit $(D-2)$-sphere, G_D is Newton's constant in D dimensions, M is the initial mass of each wormhole mouth, ℓ is the AdS length, and r_H is the initial horizon radius, defined such that $f(r_\mathrm{H}) = 0$. For $r > r_\mathrm{H}$, $f(r)$ is strictly positive, so r^* and X are monotonic in r. As we move from B_i at $X = 0$ towards $\tilde{L}_0$ or $\tilde{R}_0$ at X_L and X_R, the area of the cross-section of $\Sigma_0 \cap W_{t^*}$ for the surface parameterized by $X(\phi)$ [or equivalently $r(\phi)$], for $(D-2)$ angular variables ϕ, attains its minimum at B_i, where $r(\phi)$ is identically r_H, its minimum on $\Sigma_0 \cap W_{t^*}$. $\square$

We now turn to the behavior of the cross-sectional area of $M_+ \cap \dot{J}^-[\Sigma_{t^*}\backslash W_{t^*}]$.

Proposition 3. The cross-sectional area of $M_+ \cap \dot{J}^-[\Sigma_{t^*}\backslash W_{t^*}]$ is nondecreasing towards the future.

Proof. Let us label the left and right halves of $M_+ \cap \dot{J}^-[\Sigma_{t^*}\backslash W_{t^*}]$ as $\tilde{L}$ and $\tilde{R}$, so the boundary of $\tilde{L}$ is just $\tilde{L}_0 \cup L_{t^*}$ and similarly for $\tilde{R}$. We note that both $\tilde{L}$ and $\tilde{R}$ are generated by outgoing null geodesics. Suppose that some segment of $M_+ \cap \dot{J}^-[\Sigma_{t^*}\backslash W_{t^*}]$ has area decreasing towards the future. We can without loss of generality restrict to the left null surface, which we then assume has decreasing area along some segment.

We first observe that since the apparent horizons are null or spacelike and since $\tilde{L}$ is part of the null boundary of the past of a slice through the outermost apparent horizon, all outer trapped surfaces must lie strictly inside $\tilde{L} \cap \Sigma_t$ for all spacelike slices Σ_t for $t \in [0, t^*]$.

Let us define an affine parameter $\tilde{\lambda}$ for $\tilde{L}$, for which $\tilde{\lambda} = 0$ on $\tilde{L}_0$ and $\tilde{\lambda} = 1$ on L_{t^*}, and consider the expansion $\tilde{\theta} = \nabla_\mu \tilde{k}^\mu$, where $\tilde{k}^\mu = (\mathrm{d}/\mathrm{d}\tilde{\lambda})^\mu$. In order for the area to be strictly decreasing, there must be some open set U for which $\tilde{\theta}(\tilde{\lambda}) < 0$ for

$\tilde{\lambda} \in U$. By continuity of the spacetime, there must exist $\tilde{t}$, where we can choose the affine parameterization such that $\Sigma_{\tilde{t}} \supset \tilde{L}(\tilde{\lambda})$ for some $\tilde{\lambda} \in U$, such that $\Sigma_{\tilde{t}}$ contains a region $V \supset \tilde{L}(\tilde{\lambda})$ for which $\tilde{\theta} \leq 0$ for all outgoing orthogonal null congruences originating from V. Then V is an outer trapped surface not strictly inside $\tilde{L} \cap \Sigma_{\tilde{t}}$. This contradiction completes the proof. $\qquad\square$

Thus, we have constructed a lower bound for the area of C_t.

Proposition 4. The area of C_t is lower bounded by the sum of the areas of the initial bifurcation surfaces B_i.

Proof. To prove a lower bound on the maximin area, $\text{area}[C_{t*}]$, it suffices to exhibit an achronal surface through W_{t*} for which the minimal cross-sectional area is equal to the desired lower bound. Such a surface is given by $\Gamma_{t*}(0)$ in Eq. (5.4.8): by Proposition 2, $\sum_i \text{area}[B_i]$ is the minimal cross-sectional area of $\Sigma_0 \cap W_{t*}$ and, in particular, $\sum_i \text{area}[B_i] \leq \text{area}[\tilde{L}_0] + \text{area}[\tilde{R}_0]$. By Proposition 3, the minimal cross-sectional area of $M_+ \cap \dot{J}^-[\Sigma_{t*} \backslash W_{t*}]$ is $\text{area}[\tilde{L}_0] + \text{area}[\tilde{R}_0]$. Thus, $\Gamma_{t*}(0)$ is an achronal slice through W_{t*} with minimal cross-sectional area equal to $\sum_i \text{area}[B_i]$. $\qquad\square$

Finally, as an immediate corollary, we have the gravity dual of entanglement conservation.

Theorem 1. *For the family of spacetime regions W_t defined as the causal diamonds anchored on the piecewise-connected outermost apparent horizons L_t and R_t for an arbitrary set of dynamical, past-initialized wormholes and black holes satisfying the null curvature condition, the corresponding maximin surface C_t dividing the left and right collections of wormholes has an area independent of t, equaling the sum of the areas of the initial bifurcation surfaces for the wormholes linking the left and right sets of horizons.*

Proof. By Proposition 1, $\text{area}[C_t] \leq \sum_i \text{area}[B_i]$, while by Proposition 4, $\text{area}[C_t] \geq \sum_i \text{area}[B_i]$. Hence,

$$\text{area}[C_t] = \sum_i \text{area}[B_i]. \qquad (5.4.11)$$

$\qquad\square$

Thus, the maximin surface dividing one collection of wormhole mouths from another has an area that is conserved under arbitrary spacetime evolution and horizon mergers as well as arbitrary addition of matter satisfying the null energy condition.

Viewing the maximin surface area as the entanglement entropy associated with the left and right sets of horizons in accordance with the HRT prescription, we have proven a statement in general relativity that is a precise analogue of the statement in Sec. 5.2 of conservation of entanglement under evolution of a state with a tensor product unitary operator.

5.5 Conclusions

The proposed ER=EPR correspondence is surprising insofar as it identifies a generic feature (entanglement) of any quantum mechanical theory with a specific geometric and topological structure (wormholes) in a specific theory with both gravity and spacetime (quantum gravity). Until an understanding is reached of the geometrical nature of the "quantum wormholes" that should be dual to, e.g., individual entangled qubits, it will be difficult to directly establish the validity of the ER=EPR correspondence as a general statement about quantum gravity. In a special limiting case of quantum gravity—namely, the classical limit, which gives general relativity—this task is more tractable. In this paper, we have provided a general and explicit elucidation of the ER=EPR correspondence in this limit. For a spacetime geometry with an arbitrary set of wormholes and black holes, we have constructed the maximin area of the multi-wormhole throat separating a subset of the wormholes from the rest of the geometry, the analogue of the entanglement entropy of a reduced density matrix constructed from a subset of the degrees of freedom of a quantum mechanical state. We then proved that the maximin area is unchanged under all operations that preserve the relation between the subset and the rest of the geometry, the equivalent of quantum mechanical operations that leave the entanglement entropy invariant. We have therefore completely characterized the ER=EPR relation in the general relativistic limit: the entanglement entropy and area (in the sense defined above) of wormholes obey precisely the same rules.

In addition to providing an examination of the ER=EPR duality, our result constitutes a new area theorem within general relativity. The maximin area of the wormhole throat is invariant under dynamical spacetime evolution and the addition of classical matter satisfying the null energy condition. The dynamics of wormhole evolution were already constrained topologically (see Ref. [130] and references therein), but this result goes further by constraining them geometrically. Note that throughout this paper we have worked in asymptotically AdS spacetimes in order to relate our results to a boundary theory using the language of the AdS/CFT corre-

spondence, but our area theorem is independent of this asymptotic choice provided that all of the black holes are smaller than the asymptotic curvature scale.

In the classical limit, we have characterized and checked the consistency of the ER=EPR correspondence in generality. However, extending these insights to a well-defined notion of quantum spacetime geometry and topology remains a formidable task. Understanding the nature of the ER=EPR duality for fully quantum mechanical systems suggests a route toward addressing the broader question of the relationship between entanglement and geometry.

Acknowledgments

Rescuing Complementarity with Little Drama

6.1   Introduction

The information paradox [260] and its more modern AMPS incarnation [261, 262] are deeply puzzling issues lying at the center of any attempts at reconciling quantum mechanics with gravity. Black hole complementarity, as proposed by [263], attempted to resolve the information paradox by asserting that information that falls into the black hole interior is also retained at the stretched horizon. Observers are only able to access this information in one of two "complementary" descriptions, either in the interior or at the horizon, so that the apparent violation of the no-cloning theorem visible in a global description could never be verified. AMPS, however, considered a scenario in which an observer first collects information on the outside by gathering Hawking radiation, then jumps through the horizon and into the black hole interior. Assuming standard postulates of black hole complementarity, namely

1. unitarity,

2. the validity of low-energy effective field theory outside the stretched horizon,

3. that the black hole is a quantum mechanical system with dimension given by $e^{A/4}$,

and further

4. that the horizon is not a special place—that "no drama" happens at the horizon, so an observer can actually enter the black hole interior,

AMPS pointed out an apparent violation of monogamy of entanglement[1] among three systems: the black hole interior, the recently emitted Hawking radiation (late radiation), and the previously emitted Hawking radiation (early radiation). To avoid this violation, it therefore seemed necessary to give up one of the assumptions mentioned above, all of which are cherished pillars of modern physics. Giving up the final assumption would mean that observers who attempt to enter the black hole would be violently destroyed by high-energy excitations, hence the name "firewall paradox."

This led to a flurry of attempts to resolve the paradox by weakening one or more of the core axioms, or by changing the paradigm completely [264–273]. Reaching consensus as to which resolution is the correct one has proven challenging.

---

[1]Monogamy of entanglement is the statement that no single qubit can be simultaneously maximally entangled with two different systems.

An interesting proposed resolution to the information paradox, based on arguments from computational complexity, was given by Harlow and Hayden [274]. They argued that the part of the AMPS experiment where the experimenter has to decode[2] entanglement between the old radiation and the late radiation of the black hole involves an extremely difficult computational task. Under very plausible conjectures in computational complexity[3], the time required to perform this quantum computation in general would be exponentially longer than the evaporation time of the black hole. Thus, by the time that the entanglement is decoded, there will remain no black hole within which to check for the violation of monogamy of entanglement. While the two quantum mechanical descriptions of the black hole appear to imply a violation of monogamy, this apparent violation cannot be "revealed" by the AMPS experiment, and thus the experimenter does not see any contradiction with quantum mechanics. Just like the original violation of no-cloning in black hole complementarity itself, this would signal that only the various partial descriptions accessible by a single observer should be considered.

The main appeal of this argument is that it does not require a weakening of any of the core assumptions mentioned previously. However, it is not without its vulnerabilities. For example, Oppenheim and Unruh [276] gave an argument showing that a very motivated experimenter could evade the Harlow-Hayden complexity barrier by offloading the hard computation into a "precomputation" phase before the black hole had even formed, and then perform the AMPS experiment efficiently using the "cached computation." Another vulnerability is that the computational hardness of the Harlow-Hayden argument assumes that the black hole in question somehow encodes a cryptographically difficult one-way function; however, one may be able to set up a black hole so that the entanglement decoding task is particularly easy [277].

Nevertheless, the Harlow-Hayden proposal remains a compelling one, and it sets the context for the argument that we present in this paper. Here, we also study whether ideas from information theory and computer science can help resolve the information paradox, but in another setting: whereas Harlow and Hayden focus on the computational complexity of the AMPS experiment *outside* the black hole, we examine the information processing that must be performed *inside* the black hole in order to check for violations of monogamy of entanglement. This is a potentially

---

[2]To "decode the entanglement" of a state $|\psi\rangle_{AB}$ is to act with local unitaries on A and B to create a Bell pair across A and B. This is similar to the notion of entanglement distillation [275], except here we have only one copy of the state $|\psi\rangle_{AB}$, whereas in distillation one has multiple identical copies of the state.

[3]Namely, that quantum computers cannot efficiently invert *cryptographic one-way functions*.

different line of argument, because while it might be possible to evade computational limits outside of the horizon [262, 276], one certainly cannot extend one's time inside the horizon, as an infalling observer invariably hits the singularity in a bounded amount of time.

In this paper we study an observer who begins outside of an evaporating Schwarzschild black hole well after the Page time and who has learned that a subset of late Hawking radiation that she holds is maximally entangled with the early Hawking radiation[4]. We suppose that the observer then enters the black hole, sees no firewall, and then attempts to decode maximal entanglement between the late radiation that she holds and the black hole interior. If she succeeds in completing this task, she can then perform measurements on an ensemble of her decoded Bell pairs in order to probabilistically detect a violation of monogamy of entanglement. We compare the proper time it takes for the observer to perform this procedure with the infall time before the observer hits the singularity. We find that, under the assumption that the subsystem of the black hole interior with which the observer's late radiation is entangled has diffused throughout the whole interior at the time she crosses the horizon, the observer will not have enough time to complete even the first step of the procedure, i.e., entanglement decoding, before encountering the singularity. As such, while a global description, if it existed, would contain an implicit violation of monogamy of entanglement, an observer who entered the black hole would unable to directly verify any such violation. Therefore, our resolution of the firewalls paradox is similar in spirit to complementarity [263] in the sense that apparent global violations of quantum mechanics are not verifiable by local observers.

The assumption that we make about dynamics inside the horizon is a mild weakening of the no-drama condition that is typically considered: while we expect no-drama to hold for macroscopic, classical objects that cross the event horizon, fine-grained quantum information should be *scrambled* throughout the black hole's degrees of freedom, regardless of whether these degrees of freedom are described as the black hole horizon or as the black hole interior. In particular, the assertion that an observer inside the black hole sees such scrambling is the novel assumption of our paper. We thus call this assumption "little drama," and it is central to our argument.

The organization of this paper is as follows. In Section 6.2 we review facts about black holes and their scrambling from the perspective of different observers in spacetime. In Section 6.3, we focus on the specific task of collecting a late-time

---

[4]Though this is the task that Harlow and Hayden argue is difficult, we assume for the purpose of the argument that this task has been achieved.

Hawking radiation particle, assess the degree of scrambling that has occurred prior to the observer crossing the stretched horizon of the black hole, and give a discussion of the little-drama condition. In Section 6.4, we combine all the ingredients from the previous sections and analyze the time needed to perform the task of checking for violations of monogamy. Finally, we discuss and conclude in Sections 6.5 and 6.6.

6.2 Background: Black Holes and Scrambling

In the thought experiments to follow, we will consider black holes that formed from the gravitational collapse of matter and that eventually evaporate into a gas of Hawking radiation. We will assume that the initial mass of any black hole that we consider is large enough that physics outside the black hole is well-described by effective field theory on a black hole background in regions of spacetime that are sufficiently distant from the end of evaporation. We will also suppose that the process of black hole formation and evaporation is a fundamentally unitary process. As such, if the matter that collapsed to form a black hole was initially in a pure quantum state, then the state of the Hawking radiation after evaporation—as well as any combined intermediate state of the black hole and hitherto emitted Hawking radiation—is also a pure state.

Consider now some observer who resides outside the black hole. We will adopt the viewpoint that such an observer's observations are described according to complementarity [263] and the membrane paradigm [278]. Explicitly, suppose that the black hole spacetime is foliated by some set of achronal (spacelike or null) surfaces with respect to which the observer performs field-theoretic calculations. In accordance with complementarity, an observer outside the black hole should not associate a Hilbert space to an entire surface Σ if it intersects the event horizon. In such a case, she instead organizes the physical Hilbert space associated to Σ into a tensor product $\mathcal{H} = O \otimes D$. The space O describes the degrees of freedom on the portion of Σ that lies outside of the black hole, and D is a Hilbert space that describes the black hole's degrees of freedom and that is localized about the event horizon (Fig. 6.1). From the outside observer's point of view, all of physics is described by, and all processes play out in, these two Hilbert spaces; she never has to (and in fact *may not*) make reference to the the black hole interior.[5]

---

[5]See also [279] (in particular Sec. 4) as well as Sec. 6.5.4 for further discussion of the way in which $\mathcal{H}$ factorizes and the ways in which different factorizations are related as a consequence of assuming complementarity.
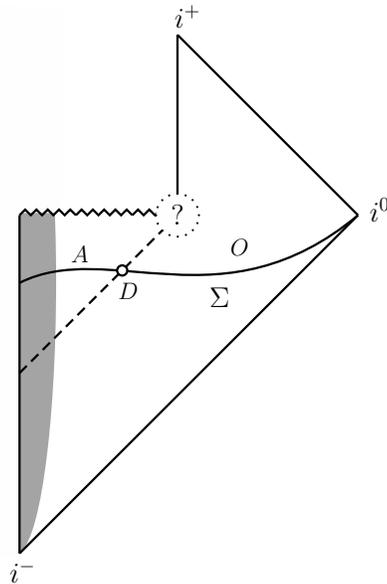
Figure 6.1. Penrose diagram of a black hole that forms from the gravitational collapse of matter and that ultimately evaporates.

We will suppose that D is localized to the stretched horizon of the black hole [263]. We take the outer boundary of the stretched horizon to be at a proper distance on the order of a Planck length above the event horizon. As such, the outer boundary of the stretched horizon is a timelike surface with which an outside observer can interact.

Despite the fact that a complete theory of quantum gravity is not known and that the full dynamics of black holes are not understood, it is widely expected that the quantum state of matter gets scrambled when it enters the stretched horizon [280–282]. There are many possible ways to define scrambling, but informally speaking, a system scrambles if it diffuses quantum information over all its degrees of freedom. In particular, a black hole has scrambled the information in a small subset $D' \subset D$ when any initial entanglement between D' and the outside O gets distributed evenly throughout D, i.e., when almost all small subsets of D have nearly the same amount of entanglement with O. After scrambling, an observer cannot recover this entanglement unless she examines a sizable fraction of the entire horizon D.

The characteristic timescale over which scrambling occurs, called the scrambling time, is given by

$$t_s = \frac{1}{2\pi T} \ln S \,, \tag{6.2.1}$$

where T and S are the temperature and entropy of the black hole respectively [280, 281, 283–285]. (Both in this expression and throughout the paper we have set $c = k_{\mathrm{B}} = \hbar = 1$.) This time is measured with respect to the clock of an asymptotic observer who is far away from the black hole. For example, for a Schwarzschild black hole in $3 + 1$ dimensions, the metric is given by

$$ds^2 = -\left(1 - \frac{r_s}{r}\right) dt^2 + \left(1 - \frac{r_s}{r}\right)^{-1} dr^2 + r^2\, d\Omega_2^2, \tag{6.2.2}$$

the temperature is

$$T = \frac{1}{8\pi G M} = \frac{1}{4\pi r_s}, \tag{6.2.3}$$

and the entropy is

$$S = \frac{A}{4G} = \frac{4\pi r_s{}^2}{4 l_P^2} = \frac{\pi r_s{}^2}{l_P^2}. \tag{6.2.4}$$

As such, the scrambling time is given by

$$t_s = r_s \ln \frac{\sqrt{\pi} r_s}{l_P}. \tag{6.2.5}$$

The event horizon is located at $r = r_s = 2GM$, and l_P denotes the Planck length. Importantly, a stationary observer who hovers at some fixed value of $r = r_0$ above the black hole sees scrambling happen faster, since her clock ticks faster relative to Schwarzschild time. In other words, the scrambling time as measured in the proper time of a stationary observer at coordinate height r_0 is

$$\tau_s(r_0) = \sqrt{1 - \frac{r_s}{r_0}}\, t_s. \tag{6.2.6}$$

In particular, we can work out what the scrambling time at the stretched horizon must be. If we fix the boundary of the stretched horizon to lie at a proper distance l_P above the event horizon, one finds that this corresponds to a coordinate distance $r = r_s + \delta r$, where

$$\delta r = \frac{l_P^2}{4 r_s} + O\left(\frac{l_P^3}{r_s{}^2}\right). \tag{6.2.7}$$

It then follows that

$$\tau_s(r_s + \delta r) = \sqrt{\frac{l_P^2}{l_P^2 + 4r_s{}^2} r_s} \ln\left[\frac{\sqrt{\pi}r_s}{l_P}\right]$$

$$\approx \frac{l_P}{2} \ln\left[\frac{\sqrt{\pi}r_s}{l_P}\right] \, , \tag{6.2.8}$$

which is consistent with other calculations of the scrambling time at the stretched horizon [280, 281].

6.3 Hawking Radiation and Scrambling: What Alice Sees

Having established the preliminaries, we can now begin to investigate the central question of this work: whether an observer who crosses the event horizon of an evaporating black hole can, in the absence of a firewall, verify a violation of monogamy of entanglement before she hits the singularity. The answer to this question depends on several considerations: in particular, the nature of scrambling from the point of view of an observer inside the black hole, under what circumstance an ingoing Hawking mode is scrambled before an observer carrying the corresponding outgoing mode crosses the horizon, and the difficulty of undoing scrambling inside the black hole. We address the first two points, the nature of scrambling and under what conditions scrambling occurs, in this section. In particular, we motivate the little-drama assumption used in the argument of this paper.

6.3.1 Scrambling, inside and out

Suppose that Alice has been monitoring a black hole since its formation and that she collects any Hawking radiation that it emits. At some point well past the Page time, she decides to perform her ultimate experiment: an experimental test of the AMPS paradox. To this end, she collects k particles of (late) Hawking radiation and first checks whether they are maximally entangled with the radiation that was emitted earlier. Let us momentarily grant Alice unlimited computational power outside of the black hole and suppose that she finds that these late quanta of radiation are indeed maximally entangled with the early radiation. She then holds on to these final Hawking particles and enters the black hole. To her transient relief, suppose that she does not encounter a firewall at the horizon. As such, suspecting a possible violation of monogamy of entanglement, her next objective is to check whether the

k Hawking particles that she collected outside of the black hole are entangled with degrees of freedom in the black hole interior.

Recall that Hawking radiation consists of paired entangled excitations of field modes. The outgoing modes constitute the radiation that is visible to stationary observers, but for each outgoing mode there is also an ingoing mode which remains confined to the black hole interior. In principle, Alice's task is to "catch up" with the ingoing excitations that correspond to the k particles that she collected and check whether they are entangled. In the next section, we will consider whether and how Alice can actually perform this check. For now, we will consider a prerequisite question: what do the ingoing excitations look like to Alice should she catch up to them inside the black hole?

Because of complementarity, while Alice is outside of the black hole, she should not think of an ingoing excitation as some particle which falls toward the singularity. Rather, she sees it as some excitation of the stretched horizon, which begins to scramble as the dynamics of the stretched horizon unfold. Yet, also because of complementarity, Alice's description of physical processes changes once she crosses the event horizon of the black hole. The stretched horizon is no more and she is now fully entitled to describe physics in the black hole interior. For example, she can now associate a Hilbert space with each of her past lightcones and make the division $\mathcal{H} = A \otimes O$, where A and O describe degrees of freedom on the intersection of her past lightcone with the interior and exterior of the black hole, respectively. It is in this frame that she must look for the ingoing excitations.

Our aim is to understand the interplay between scrambling in the stretched horizon and the change in Alice's description of physics as she enters the black hole. Or, in other words, complementarity maintains that physics as described from inside and outside the black hole should, in an appropriate sense, be equivalent; we want to understand how scrambling—which is a process that occurs from an outside observer's point of view—appears to an observer inside the black hole.

To be more precise, suppose that Alice follows a timelike trajectory $\mathcal{A}$ that crosses the event horizon and ultimately hits the singularity, as shown in Fig. 6.2. (Partially) foliate the spacetime with her past lightcones. When she is inside the black hole, we associate A to the portion of her lightcone that lies inside the black hole. For all of her lightcones, we associate O to the part of the lightcone that lies outside the black hole and D to the surface where her lightcone intersects the stretched horizon. According to complementarity, we postulate that for each lightcone whose tip lies
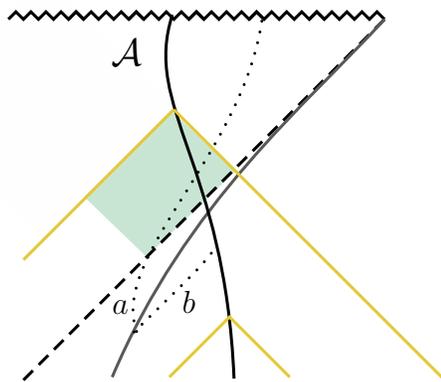
Figure 6.2. Alice's trajectory $\mathcal{A}$ and past lightcones (shown in yellow) as she falls toward the singularity. The stretched horizon is shown in grey, and the trajectories of the outgoing and ingoing Hawking particles are shown as dotted lines. We suggest that scrambling causes information about the ingoing excitation to spread out behind the event horizon so that it is delocalized on the intersection of Alice's past lightcones with the causal future of the excitation's horizon crossing point (shaded region).

inside the black hole, there exists a unitary map

$$\mathcal{U}_{\mathrm{comp}} : D \otimes O \longrightarrow A \otimes O \qquad (6.3.1)$$

that relates the complementary descriptions of physics on either side of the event horizon. ($\mathcal{U}_{\mathrm{comp}}$ is a effectively a change of basis.) If scrambling amounts to a unitary process in the stretched horizon, $U_{\mathrm{scr}} : D \to D$, then scrambling causes the state of the ingoing modes that Alice finds inside the black hole to evolve according to the action of

$$\tilde{U}_{\mathrm{scr}} \equiv \mathcal{U}_{\mathrm{comp}} \left(U_{\mathrm{scr}} \otimes I_{\mathrm{out}} \right) \mathcal{U}_{\mathrm{comp}}^{\dagger} . \qquad (6.3.2)$$

Intuitively, one would expect that scrambling should persist behind the event horizon. For instance, if one were to drop a qubit into the stretched horizon and wait for it to be well-scrambled, it would be surprising to find it more or less intact and localized after jumping into the black hole. Moreover, such a discovery would be troubling in light of Hayden and Preskill's finding that the information contained in that qubit is very rapidly returned to the exterior of the black hole [280]. Mathematically, this expectation is equivalent to the statement that we do not expect the unitary operator (6.3.2) to act trivially on the physically relevant states in A. We note, however, that it is not logically impossible that $\mathcal{U}_{\mathrm{comp}}$ exactly undoes the action

of U_{scr}.

On the other hand, it would also be desirable to reconcile the unitary (6.3.2) with the semiclassical expectation that spacetime and macroscopic gravitating objects near the event horizon are well-described by general relativity. Put another way, the field equations of general relativity should be sufficient, at least to a first approximation, to track classical matter thrown into the black hole on timescales where Hawking evaporation is unimportant. For example, from a semiclassical point of view, if you were to drop a rock into a black hole, you would still expect to find the rock on its freefall trajectory if you accelerated to catch up with it behind the event horizon.

We therefore expect that $\tilde{U}_{\mathrm{scr}}$ should act highly nontrivially on fine-grained quantum degrees of freedom, but preserve the coarse-grained state of macroscopically robust and decohered objects. More precisely, we expect that the classical geometry inside the black hole should be described by some coarse-graining of A, and that the resulting coarse-graining of $\tilde{U}_{\mathrm{scr}}$ should act trivially on classical states in this reduced Hilbert space, but that its action on typical states in the full Hilbert space is highly nontrivial. In particular, this implies that typical ingoing Hawking quanta, which are of course fully quantum, should be rapidly mixed with the rest of the modes in the black hole interior. On the other hand, a classical observer like Alice should be relatively unaffected by the same dynamics, though of course she will be destroyed in an infall time anyway. We leave it as an open problem to find a reasonable family of scrambling unitaries that implements little-drama: i.e., dynamics that scrambles small quanta, but leaves classical objects largely intact. However the arguments that follow will only make use of the fact that the ingoing Hawking quanta are rapidly scrambled over the black hole interior, and not the fact that macroscopic objects are preserved. As such, we will model U_{scr} (and hence $\tilde{U}_{\mathrm{scr}}$) as a generic unitary[6].

We emphasize that the dynamics that we have proposed constitute a violation of the no-drama condition, albeit a far milder one than firewalls. In classical general relativity, the equivalence principle remains intact: the black hole geometry is still described by the Schwarzschild metric, and nothing special happens at the horizon. Even semiclassically, expectation values of operators should remain unchanged: we are not changing the emission rate of Hawking quanta or the effective temperature of the black hole. However, working with Hawking emission on a particle-by-particle basis requires a more detailed description. We can write the quantum state describing the evaporating black hole in a basis of states which each contain Hawking parti-

---

[6]See Sec. 6.5 for a discussion on this simplifying assumption.

cles. In each basis state, individual Hawking quanta are pair-produced as genuine particles (i.e., wavepackets) at a specific spot on the horizon of the black hole, with one wavepacket excitation describing a particle produced in A and a corresponding particle in O. In each basis state, $\tilde{U}_{\text{scr}}$ acts to rapidly spread the excitation in A into many other modes, so that after a scrambling time it can no longer be described as a wavepacket or particle. It is this evolution, which differs dramatically from the propagation of a particle on an empty background metric, that can be seen as violating no drama.

6.3.2   Scrambling and kinematics

Next we investigate under what circumstances scrambling of the ingoing modes occurs from Alice's point of view. Let a clock fixed at the stretched horizon begin ticking when Alice's final particle of Hawking radiation is emitted. We shall use its reading when Alice reaches the stretched horizon to determine whether or not the corresponding ingoing excitation—which, again, Alice sees as an excitation on the stretched horizon while outside the black hole—has scrambled.

In principle, Alice could wait arbitrarily closely to the stretched horizon so that the ingoing excitation has little time to scramble. We note, however, that the scrambling time at the stretched horizon is a fantastically small amount of time. For example, for a supermassive black hole like Sagittarius A* with a mass of about four million solar masses, Eq. (6.2.8) predicts that the scrambling time at the stretched horizon should be $\tau_s \approx 3 \times 10^{-42}$ s, or about 50 Planck times. As such, Alice does not have much time at all outside of the black hole before scrambling happens, and in practice she will have some amount of computational overhead if she verifies the entanglement between late radiation and early radiation before entering the black hole. Furthermore, if Alice collects $k > 1$ Hawking particles, then scrambling of the first $k - 1$ ingoing excitations is virtually guaranteed to have happened before Alice can cross the horizon. This is because the average rate of Hawking emissions is (much) slower than the rate of scrambling [286, 287]. Consequently, instances where Alice can cross the horizon before ingoing modes have scrambled are $(k-1)$-fold exponentially suppressed.[7] As we will discuss in the next section, Alice will need to collect

---

[7]From [286], the cumulative Hawking emission rate for a Schwarzschild black hole is about 10^{-4} c^3/GM, so take the characteristic timescale of Hawking emissions to be $t_H \sim 10^4$ GM/c^3. Note that this is measured in Schwarzschild time, so with the relevant boost factor of $l_P/2r_s$ and for the supermassive black hole discussed above, the characteristic (proper) timescale of Hawking emissions at the stretched horizon is about $(10^3 - 10^4)$ l_P/c, which is much larger than the scrambling time. Also c.f. footnote 9 below.
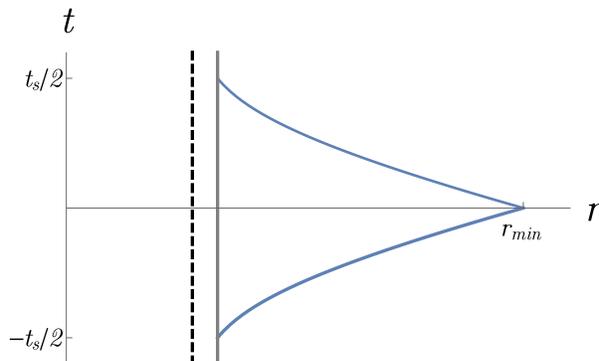
Figure 6.3. Minimum height above which scrambling is guaranteed to occur.

$k > 1$ Hawking particles in order to be statistically confident in her measurements inside the horizon.

Separately from the considerations above, it is also interesting to ask what is the theoretical minimum height at which Alice can wait above the black hole is above which scrambling is guaranteed to have happened when Alice enters the black hole. This is the height for which exactly one scrambling time elapses at the stretched horizon in the time it takes a light ray to make a round trip between the stretched horizon and a mirror at the height in question. This situation is depicted in Schwarzschild coordinates in Fig. 6.3.

The radial lightlike geodesics are given by

$$ r - (r_s + \delta r) + r_s \ln \left[\frac{r_s - r}{r_s - (r_s + \delta r)} \right] = \pm \left(t + \frac{t_s}{2} \right) , \qquad (6.3.3) $$

with t_s and δr as given in Eqs. (6.2.5) and (6.2.7) respectively. The minimum coordinate height is obtained by setting $t = 0$ in Eq. (6.3.3) and solving for r:

$$ r_{\min} = r_s \cdot W \left(\frac{\delta r}{r_s} \exp \left[\frac{2\delta r + t_s}{2r_s} \right] \right) . \qquad (6.3.4) $$

In the above, $W(\cdot)$ denotes the Lambert W function. The minimum proper distance

is therefore given by

$$
\begin{aligned}
\tilde{r}_{\min} &= \int_{r_s}^{r_{\min}} \left(1 - \frac{r_s}{r}\right)^{-1} \, dr \\
&= 2\sqrt{r_s e^{t_s/2R}} \sqrt{\delta r} + O\left((\delta r)^{3/2}\right) \\
&\approx \sqrt{\pi} r_s \, .
\end{aligned}
\tag{6.3.5}
$$

This result is interesting in light of proposals by Nomura, Sanches, and Weinberg [272] and by Giddings [288] which both suggest that Hawking radiation is largely invisible to observers unless they are at least on the order of a few Schwarzschild radii away from the horizon of a black hole, which further limits Alice's ability to evade scrambling.

6.4 Computation behind the Horizon

To summarize the previous section, if excitations at the stretched horizon are scrambled when Alice reaches the stretched horizon, then we are proposing that the state of the ingoing Hawking modes is thoroughly mixed with other degrees of freedom in the black hole's interior. In this section we assume that this scrambling has had time to occur; as we explain in Sec. 6.3.2, such a situation should be generic. As such, Alice is forced to access and process a large number of degrees of freedom that are distributed throughout the interior of the black hole if she wants to verify monogamy of entanglement. In this section, we discuss how to model the task of verifying entanglement and we investigate its complexity. In the rest of this paper we will set $l_P = 1$ for brevity.

6.4.1 Model for verifying entanglement

Following the convention of [261], we continue to denote the Hilbert space of the interior of the black hole by A, and we label the Hilbert spaces of the early radiation and late radiation by R and B respectively (so that R and B are subsets of the space O that we defined in Sec. 6.2). Let $b^{(k)} \subset B$ denote the Hilbert space of the k outgoing Hawking modes that Alice collected and $a^{(k)} \subset A$ the Hilbert space of the corresponding k ingoing modes. We model $b^{(k)}$ and $a^{(k)}$ each as a collection of k qubits. Referring to Eq. (6.3.1), since the Hilbert space O is the same in both

complementary descriptions of physics[8], it follows that $|A| = |D| = e^{S_{BH}}$, where S_{BH} is the Bekenstein-Hawking entropy of the black hole and where $|\cdot|$ denotes the dimension of a Hilbert space. As such, we model A as a collection of $n \sim S_{BH}$ qubits that are distributed throughout the interior of the black hole and that are visible to Alice on her past lightcones.

First, what do we mean by "detecting a violation of the monogamy of entanglement?" This is nonsensical from the point of view of quantum mechanics, in which monogamy of entanglement is inviolable. Here, we are given an *apparent* quantum description of entanglement between $b^{(k)}$ and R outside the horizon, and an *apparent* quantum description of entanglement between $b^{(k)}$ and $a^{(k)}$ across the horizon. While the AMPS paradox shows that there cannot be a global quantum picture that is consistent with both descriptions, the crucial question now is whether Alice can perform an experiment to detect this paradox: in other words, whether she can verify the entanglement between R and $b^{(k)}$, and then verify the entanglement between $b^{(k)}$ and $a^{(k)}$. If Alice succeeds in verifying both entanglements, then we say that she has detected a violation of monogamy.

What do we mean by verifying entanglement? In quantum theory, there is no measurement that reliably distinguishes between entangled states and unentangled states—this is because the set of unentangled pure states is non-convex. However, it is possible to *statistically* test if an unknown state is in a *particular* entangled state. For example, if we let $|\Phi\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ denote an EPR pair, then the two-outcome measurement $M = \{|\Phi\rangle\langle\Phi|, I - |\Phi\rangle\langle\Phi|\}$ will probabilistically indicate whether a given pair of particles $|\psi\rangle$ is an EPR pair or not. If $|\psi\rangle$ is indeed an EPR pair, then this measurement will always return outcome $|\Phi\rangle\langle\Phi|$ with certainty. On the other hand, if $|\psi\rangle$ is an unentangled state $|\phi\rangle \otimes |\theta\rangle$, then it will return outcome $I - |\Phi\rangle\langle\Phi|$ with probability at least $1/2$. While the error of this statistical test is rather large, it can be reduced exponentially by repeating it many times. Let V and W denote two disjointed quantum systems. When we say that Alice has "verified maximal entanglement between V and W," we mean that Alice has decoded k pairs of particles from V and W, measured each pair using the two outcome measurement M, and verified that all k pairs projected to an EPR pair. This occurs with probability 1 if Alice did indeed decode k EPR pairs; if V and W were unentangled, then this occurs with probability at most 2^{-k}. Therefore as k grows, the probability that Alice thinks that V and W are entangled (when they are actually unentangled) becomes exponentially small. For example, if Alice wants to obtain 5 sigma certainty (error

---

[8]We stress, though, that $\mathcal{U}_{\mathrm{comp}}$ does not factorize over D and O.

probability 1 in 3.5 million) that V and W share maximally entangled particles, she only needs to decode $k = 22$ EPR pairs from V and W.

6.4.2 Alice's computational task

In this argument, we focus on Alice's task of verifying the entanglement between $b^{(k)}$ and $a^{(k)}$ when she jumps into the black hole—we will assume that she has already verified the entanglement between $b^{(k)}$ and R prior to jumping in. We consider the quantum description of the black hole interior A, along with the late-time Hawking modes $b^{(k)}a^{(k)}$. Consider the moment at the stretched horizon that k Hawking pairs $b^{(k)}a^{(k)}$ were produced[9]. The state of the Hawking pairs and the black hole interior can be described by the density matrix

$$\sigma^{b^{(k)}a^{(k)}A} = (|\Phi\rangle\langle\Phi|^{\otimes k})^{b^{(k)}a^{(k)}} \otimes \rho^A,$$

where $|\Phi\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ is a maximally entangled Hawking pair, and ρ^A is the density matrix of the black hole interior right before the pair production event. By Page's theorem [289, 290], after the Page time ρ^A is close to being maximally mixed; for the remainder of this argument, we will assume that ρ^A is exactly the maximally mixed state on n qubits.[10]

As discussed in the previous section, by the time that Alice arrives at the stretched horizon with $b^{(k)}$ in tow, the black hole interior (which now includes $a^{(k)}$) has experienced extensive scrambling. We model this as follows. Let U be the unitary representing the scrambling dynamics, which acts on $A' = a^{(k)}A$. From Alice's point of view, the state of the scrambled interior A' and $b^{(k)}$ can then be described by

$$\tau^{b^{(k)}A'} = (I^{b^{(k)}} \otimes U^{A'})\sigma^{b^{(k)}A'}(I^{b^{(k)}} \otimes U^{A'})^\dagger.$$

Because our understanding of the quantum mechanical evolution of black holes is rather limited, we will model the unitary U as being Haar-random. (In fact our arguments will carry through in the case that U is chosen from an ensemble of

---

[9]For simplicity here we assume that they are produced simultaneously rather than one-by-one, but this does not hinder the argument. Indeed, if they are produced sequentially, then due to arguments by Page [286, 287], the average rate of Hawking pair production is less than one pair per scrambling time. Therefore, in a sequential production picture, all but the last Hawking pair will have been scrambled by the time that Alice can enter the black hole. If Hawking radiation can be modeled thermally, sequential emission is exponentially preferred over simultaneous emission.

[10]If ρ^A is ε-close to the maximally mixed state, then our final bounds will only acquire an additional ε additive error.

efficiently constructible unitaries that is sufficiently randomizing; we will discuss this in more detail in Sec. 6.5.)

As Alice falls towards the singularity, she attempts to interact with a set S of qubits of the interior in order to recover at least one unit of entanglement between the interior and $b^{(k)}$. First, suppose S is a subsystem of A' that has at most $n/2$ qubits. Then, by [280], we have that

$$\int dU \left\| \tau^{b^{(k)}S} - \tau^{b^{(k)}} \otimes \tau^S \right\|_1^2 \leq |b^{(k)}S| \cdot \mathrm{Tr}\left[(\sigma^{b^{(k)}A'})^2\right]. \qquad (6.4.1)$$

We have that $\mathrm{Tr}\left[(\sigma^{b^{(k)}A'})^2\right] = \mathrm{Tr}\left[(|\Phi\rangle\langle\Phi|^{b^{(k)}a^{(k)}})^{\otimes k} \otimes (\rho^A)^2\right] = \mathrm{Tr}\left[(\rho^A)^2\right] = 2^{-n}$. The dimension of $b^{(k)}S$ is at most $2^{n/2+k}$, so therefore

$$\int dU \left\| \tau^{b^{(k)}S} - \tau^{b^{(k)}} \otimes \tau^S \right\|_1^2 \leq 2^{-n/2+k}.$$

Thus, by the time Alice reaches the event horizon, with probability exponentially close to one (over the choice of unitary U), any subset S of at most $n/2$ qubits of the interior of the black hole will essentially be uncorrelated with her Hawking modes $b^{(k)}$: the black hole dynamics "smears" the entanglement between $b^{(k)}$ and $a^{(k)}$ over the entirety of the black hole. This holds for as long as $k \ll n/2$, i.e., as long as the amount of material that Alice brings with her into the black hole is negligible compared to the size of the black hole[11]. Therefore, unless Alice interacts with more than half of the qubits of the black hole, she has no hope of decoding a partner qubit that is maximally entangled with $b^{(k)}$ after crossing the event horizon.

However, can Alice interact with more than half of the qubits in A'? We assume that Alice is a localized experimenter (such that she is unable to do parallel computation on a spacelike region), so that she can only process at most $O(1)$ qubits of the black hole interior per Planck time. Thus, to touch at least $n/2$ qubits, Alice would require $\Omega(n)$ Planck times. However, Alice also has no chance of doing this before experiencing an untimely demise: the longest amount of time that can elapse on Alice's clock before she reaches the singularity is $O(r_s) = O(\sqrt{n})$ in Planck units. Again, she has no hope of decoding any entanglement between $b^{(k)}$ and A'. In other words, because of black hole scrambling, Alice does not have enough time to verify the entanglement between $b^{(k)}$ and $a^{(k)}$, and thus is unable to perform the AMPS

---

[11]Otherwise, if Alice is bringing a sizable fraction of the black hole's mass with her across the horizon, this could plausibly take the state of the black hole to before the Page time, change the horizon size, or any number of other nonperturbative effects which break the setup of the paradox.

experiment.

6.5 Discussion

We now elaborate upon several aspects of our argument, including discussing possible objections.

6.5.1 Modeling scrambling dynamics

In our argument, we model the scrambling dynamics of the black hole as a generic unitary sampled from the Haar distribution. As mentioned before, we model U_{scr} as a generic unitary in order to capture the part of little-drama where fine-grained quanta get scrambled. It does not model the other part of little-drama where macroscopic objects are preserved, but we do not use this part in our argument.

An immediate objection to this modeling choice is that black hole dynamics cannot, strictly speaking, look anything like a Haar-random unitary. This is because a generic unitary will have exponential complexity: the minimum number of local quantum operations that need to be applied in order to implement the unitary – known as the *circuit complexity* of the unitary – is exponential in the number of its degrees of freedom. Assuming the Physical Church-Turing Thesis[12], an n-qubit black hole that evolves for poly(n) Planck times should only be able to realize unitaries that have poly(n) circuit complexity, where poly(n) denotes some polynomial in n. Perhaps unitary matrices with polynomial circuit complexity will not adequately "smear" entanglement across the entire black hole interior, as required by our argument.

As noted by Hayden and Preskill [280], one can model the dynamics of a black hole using *random unitary designs*. Informally speaking, unitary designs are ensembles of unitaries with polynomial circuit complexity that in many respects behave like Haar-random unitaries. In our argument, the Haar unitary ensemble can be replaced by an (approximate) unitary design and our conclusion remains essentially unchanged: unitary designs, though possessing small circuit complexity, still "smear" quantum information across all degrees of freedom. Unitary designs have been extensively studied in the quantum information community. By now, we know several examples of (approximate) unitary designs [291, 292].

---

[12]Briefly, the Physical Church-Turing Thesis states that all computations in the physical universe can be simulated, with polynomial time overhead, by a universal quantum computer.

Still, what do we mean when we say that a *particular* black hole behaves like a unitary randomly chosen from an ensemble? After all, a black hole behaves according to none other but *the* unitary given by the theory of quantum gravity. Unfortunately, since this theory is still unavailable to us, in our calculations we must make a "best guess" at what a black hole unitary must look like. Without presupposing unjustified constraints on the theory of quantum gravity, our best guess for black hole dynamics is that the Hamiltonian governing the interior should be local and strongly mixing, and that the black hole evolves in polynomial time. The Maximum Entropy Principle from statistics and learning theory tells us that our best guess for the black hole unitary is a randomly chosen one from the uniform distribution over unitaries with polynomial circuit complexity[13]. We note that this ensemble of unitaries is known to form an approximate unitary design [291], and thus has the scrambling properties required by our argument.

6.5.2 Black holes in other dimensions.

One may also object that this argument is specific to spacetimes of dimension 3+1. In higher dimensions this argument only becomes stronger, since in spacetimes with spatial dimension d, the number of qubits that make up the interior Hilbert space, $|A|$, scales like $O(r_s^{d-2})$, while the infall time scales like $O(r_s)$. As such, the infall time is increasingly smaller with respect to $|A|$ for $d > 3$. But, this is not necessarily true for lower spatial dimensions. For example, in AdS$_3$, the number of qubits and the infall time both scale linearly with r_s. Consequently, our previous trivial bound on the number of accessible qubits does not suffice here. In this case one can appeal to the fast scrambling conjecture to render the computation impossible. The fast scrambling conjecture of Sekino and Susskind [281] states (among other things) that black holes are the fastest scramblers in nature[14]. Lashkari *et al.* [282] formalized this notion in terms of quantum information by stating that black holes saturate the $r_s \log r_s$ lower bound for scrambling time. In this work, we consider a quantum complexity formulation:

---

[13]The Maximum Entropy Principle is a formalization of Occam's Razor in machine learning and statistical learning theory [293]. It says that, given a set of hypotheses consistent with one's observations, one's best hypothesis is the maximum entropy one: a randomly chosen one from that set.

[14]We note that the fast scrambling conjecture stating that the fastest scrambling time for a black hole is $r_s \log r_s$ is an asymptotic statement, and thus not broken by earlier statements of $\log r_s$ scrambling time at the stretched horizon.

Conjecture 2. Let $k \ll n/2$, i.e., let k be much smaller than the number of qubits in the black hole. Let U be the unitary corresponding to running black hole dynamics for time t on $A' = a^{(k)}A$, as measured by an asymptotic observer. Then recovering the entanglement between $a^{(k)}$ and $b^{(k)}$ from A' and $b^{(k)}$ requires time at least t. More formally, for any unitary V acting on system A', if $\nu^{b^{(k)}A'} = (I^{b^{(k)}} \otimes V^{A'}U^{A'})\sigma^{b^{(k)}A'}(I^{b^{(k)}} \otimes V^{A'}U^{A'})^\dagger$ is the state of the system after applying VU to A', and if

$$\left\|\nu^{b^{(k)}A'} - \nu^{b^{(k)}} \otimes \nu^{A'}\right\|_1^2 \geq \delta\,,$$

where δ is a small constant (say 0.01), then V has circuit depth at least t.

This is a circuit-depth version of the statement "black holes are the fastest scramblers in nature." It says that if one wishes to invert the scrambling performed by the black hole, then one requires at least the scrambling time to do so. If such a statement is true, then in our model, unscrambling the entanglement between $a^{(k)}$ and $b^{(k)}$ requires at least $r_s \log r_s$ time in any dimension, whereas the infall time scales as r_s. Therefore, such a conjecture would suffice for our arguments to hold in any dimension.

6.5.3 Localization of the experimenter.

In our argument, we assume that Alice is localized throughout our experiment, and therefore can access only $O(r_s)$ qubits after crossing the horizon. One might object that if one knew the exact dynamics of $\tilde{U}_{\text{scr}}$, one could set up the infalling matter such that a nonlocal experiment is performed on the interior modes and the result is then sent to Alice. However, this is impossible because Alice is out of causal contact with most of the black hole interior [294] from which the results of the nonlocal experiment would have to be sent. Therefore, even this non-local experiment cannot reveal entanglement between the interior and exterior Hawking modes before Alice hits the singularity.

6.5.4 Relation to prior works

We first note that in [294] arguments have already been made about the inability of the infalling observer to access the entirety of the interior of the black hole except at the singularity. These arguments are quite different in nature from the information-theoretic ones of this paper. In particular, there appears to be the possibility to work around the arguments in [294] by using multiple observers [295], something which

does not seem to be an issue in the more information-theoretic arguments of this note.

Readers may notice that our argument significantly resembles that given by Hayden and Preskill [280]. While the techniques are similar, our conclusions and assumptions differ in several ways. First, [280] concludes that black holes, rather than being information sinks, are plausibly more like information "mirrors;" information deposited into the black hole gets released (in scrambled form) as quickly as possible. On the other hand, our goal is to demonstrate a *lower bound* on Alice's ability to recover a single qubit of information within the black hole after it has been scrambled. Second, Hayden and Preskill explicitly model the joint state of the black hole and its radiation as well as some reference system as a pure state. However, in the context of the firewalls paradox, we cannot write down such a description to begin with! In our setting, we focus solely on the part of the black hole that Alice sees after she has collected her Hawking mode and has crossed the event horizon. This is consistent with complementarity; we only need to provide a valid description of physics inside the horizon, which need not be in a tensor product with the description of physics outside the horizon.

Our proposal also shares some spiritual similarities with fuzzball complementarity [296], in which undisturbed freefall through the horizon is recovered in the limit where the incident energy of the observer is much larger than the temperature of the black hole, in the sense that local properties of the infalling observer are important to consider in both cases. We note that in the context of the fuzzball program, the definition of complementarity invoked by AMPS—which we follow in Sec. 6.2 when we define the Hilbert space relevant to the problem—is replaced by a different and perhaps more correct definition involving the definition of the state along the complete slice, both inside and outside of the horizon. While it would be interesting to reformulate our results in that lens, it is perhaps unnecessary: in that limit the fuzzballs program already precludes the need for a different resolution to the information paradox! Instead, we emphasize that, even when cleaving as close to AMPS-style complementarity definitions as possible, information- and complexity-theoretic arguments by themselves strongly constrain the ability for any observer to actually observe violation of monogamy of entanglement.

We also differ from the fuzzballs approach in analyzing operationally what is possible for the observer to compute after crossing the stretched horizon of the black hole on the way to an existent singularity. In this work, the singularity plays a vital role in determining the longest possible time available to perform the computation.

But, in fuzzball complementarity, the singularity is fuzzed out and resolved at some characteristic fuzzball radius, behind which space stops existing. It may be interesting to see by what degree our bounds would tighten in the specific case of fuzzballs; we reiterate, though, that we are already able to demonstrate that we cannot operationally detect monogamy of entanglement even without the shorter longest possible time for the computation given by the fuzzball program.

Finally, we also note the recent paper [297], which provides a concrete toy model for fuzzball complementarity. It would be interesting to examine our proposals in the context of this work, since the dynamics of infalling excitations discussed in [297] may be able to inspire and inform a similarly concrete realization of the scrambling dynamics that we discussed in Sec. 6.3.1.

6.5.5 Other black hole geometries

We have thus far restricted our attention to only Schwarzschild black holes. It is a reasonable question to ask what happens once we consider other geometries with nonzero spin or charge. With regard to these, the addition of spin or charge to a black hole splits the horizon into an inner and an outer horizon. It is possible in such geometries to spend a longer amount of time between the two horizons, so in principle Alice could have enough time to complete her monogamy verification before hitting the singularity, thus implying a naive breakdown of the story up to this point. Alternatively, in maximal extensions of these black hole spacetimes, Alice could pass from the black hole interior into other asymptotically flat spacetime regions and continue to exist indefinitely.

We note, however, that the inner horizon is not entirely understood from either a general-relativistic or quantum-theoretic perspective [298, 299]. (For example, the inner horizon is strongly believed to be unstable.) As such, it is likely that our assumptions about quantum mechanics and general relativity would need to be modified (at least in the vicinity of the inner horizon) in order to discuss charged spinning black holes, and it is another question entirely what form the AMPS paradox would take if it persists.

6.6 Conclusion

We have described a resolution of the information paradox that amounts to a weakening of the no-drama condition — a new condition that we call little-drama. We suppose that quantum systems that cross the event horizon of a black hole experience

nontrivial evolution which entangles them with other degrees of freedom in the black hole interior. Such evolution inside the horizon is the complementary description of scrambling on the stretched horizon and constitutes a mild departure from the predictions of a non-gravitating field theory.

The little-drama condition allows for an apparent violation of monogamy of entanglement that is similar in spirit to the Harlow-Hayden proposal. Past the Page time, an observer can verify that early and late Hawking radiation have the right entanglement structure outside of a black hole and then smoothly pass through the event horizon. While the smooth crossing implies a violation of monogamy of entanglement—it would seem that the late radiation is maximally entangled with both the early radiation and the black hole interior—we found that the observer could not verify this violation before encountering the singularity.

It is also worth emphasizing that, as an information-theoretic proof, our arguments for larger than three spacetime dimensions are resilient to the Oppenheim-Unruh precomputation-style attacks, which are complexity-theoretic in nature. Though our complexity-theoretic argument (which holds in all dimensions) does not necessarily share this feature, it is possible that precomputation cannot simultaneously prevent both our construction and the Harlow-Hayden argument from resolving the AMPS paradox. Two distinct and mutually exclusive precomputation style attacks are required to foil both obstacles to AMPS. In the first, one collapses halves of Bell pairs into a black hole to evade Harlow-Hayden. In the second, one takes entire Bell pairs and collapses them into a black hole to evade our arguments. We note it is not simultaneously possible to do both for any single qubit. Therefore these two resolutions of the information paradox might be complementary in a different sense of the word.

Directions for future research include finding a model for black hole dynamics that faithfully captures all parts of little-drama. Other directions include working out the details for other black hole geometries with nonzero spin or charge. As previously discussed, it is not clear that such geometries would be precluded from violation of monogamy of entanglement in the same way, but a parametric comparison of how much leeway they have would be interesting to conduct. It would also be interesting if the information-theoretic proof method could be extended to spacetimes with fewer than three spatial dimensions without assuming the fast-scrambling conjecture.

De Sitter Space without Dynamical Quantum Fluctuations

7.1 Introduction

De Sitter spacetime, and approximations to it, have come to play an important role in modern cosmology, both in inflation and in the likely future evolution of our universe. The Hubble parameter in de Sitter is constant and related to the cosmological constant by $H = \sqrt{\Lambda/3}$. A stationary observer is surrounded by a cosmological horizon at a distance $R = H^{-1}$. Quantum field theory (QFT) in curved spacetime describes a unique state that is both de Sitter invariant and Hadamard (well-behaved at short distances), called the Euclidean (or Bunch-Davies [145, 146]) vacuum for a free, massive scalar field or the Hartle-Hawking vacuum [147] for an interacting scalar field. A particle detector sensitive to a field in the Hartle-Hawking vacuum will detect thermal Gibbons-Hawking radiation with a temperature $T = H/2\pi$ [148]. Each horizon-sized patch (which we will henceforth simply call a "patch") of de Sitter can be associated with an entropy equal to the area of the horizon in Planck units, $S = 3\pi/G\Lambda$. In horizon complementarity, the quantum state of the bulk of each patch can be described by a density operator defined on a Hilbert space of dimension $\dim \mathcal{H} = e^S$ [149, 150].

Conventional wisdom holds that the Hartle-Hawking vacuum experiences fluctuations, which may be thought of as either "quantum" or "thermal," since a patch is a quantum system at a fixed temperature. These fluctuations play several important roles in modern cosmological models. During inflation, when the metric is approximately de Sitter, fluctuations seed the density perturbations responsible for the cosmic microwave background (CMB) anisotropies and large-scale structure [151–153]. Eternal inflation (either stochastic [154–156] or in a landscape of vacua [157–161]) makes use of fluctuations upward in energy density, often described as "uptunneling" [162, 163]. Finally, the phenomena of Poincaré recurrences and fluctuations into Boltzmann brains can be problematic features of long-lived de Sitter phases [164–166].

We will argue that some of this conventional wisdom is wrong. Although a patch in the Hartle-Hawking vacuum is in a thermal state, we argue that it does not experience any kind of time-dependent fluctuations. The density operator in the patch takes the form $\hat{\rho} \sim e^{-\beta\hat{H}}$, where $\beta = 1/T$ and $\hat{H}$ is the static Hamiltonian. The state is stationary; there is no time dependence of any sort. While it is true that an out-of-equilibrium particle detector inside the patch would detect thermal radiation,

there are no such particle detectors floating around in the Hartle-Hawking vacuum.[1] In fact, any particle detector placed in the vacuum would equilibrate, reaching a stationary state with thermal occupation numbers [168].

Fluctuations observed in a quantum system reflect the statistical nature of measurement outcomes. Making a definite measurement requires an out-of-equilibrium, low-entropy detection apparatus that interacts with an environment to induce decoherence. Quantum variables are not equivalent to classical stochastic variables. They may behave similarly when measured repeatedly over time, in which case it is sensible to identify the nonzero variance of a quantum-mechanical observable with the physical fluctuations of a classical variable. In a truly stationary state, however, there is no process of repeated measurements and hence no fluctuations that decohere. We conclude that systems in such a state—including, in particular, the Hartle-Hawking vacuum—never fluctuate into lower-entropy states, including false vacua or configurations with Boltzmann brains.

Although our universe, today or during inflation, is of course not in the vacuum, the cosmic no-hair theorem [8–10] implies that any patch in an expanding universe with a positive cosmological constant will asymptote to the vacuum. Within QFT in curved spacetime, the Boltzmann brain problem is thus eliminated: a patch in eternal de Sitter can form only a finite (and small) number of brains on its way to the vacuum. At the same time, the standard story of inflationary perturbations remains intact: decoherence is accompanied by copious production of entropy during reheating. Our analysis of fluctuations only calls into question the idea of dynamical transitions from stationary states to states with lower entropy. We point out that the stochastic approximation in slow-roll eternal inflation [154–156] makes use of such transitions to describe putative upward fluctuations of the inflation field. Our picture rules out such fluctuations and may therefore change the conventional understanding of the conditions required for eternal inflation to occur. In particular, eternal inflation is no longer an inevitable consequence of monomial inflation potentials like $V = m^2\varphi^2$.

The cosmic no-hair theorem is given in the context of QFT in curved spacetime. Once quantum gravity is included, we need to be more careful. If we accept the notion of horizon complementarity [150, 169, 170, 263], we should not use local QFT to simultaneously describe locations separated by a horizon. Rather, we should treat each patch of eternal de Sitter space, together with its horizon, as a closed, finite-dimensional quantum system. The system is not stationary, so must

---

[1]See [167] for a discussion on the difficulties of measurements in a finite-dimensional asymptotic de Sitter space if a measuring device were indeed present.

undergo Poincaré recurrences as well as fluctuations, including into configurations we would describe as Boltzmann brains. Alternatively, there might be a higher-entropy vacuum to which the system can decay, in which case the false de Sitter vacuum patch can be thought of as an open subsystem embedded in a larger theory. If the higher-entropy vacuum is de Sitter, then the full system still has a finite-dimensional Hilbert space, subject to Poincaré recurrences and fluctuations. If there is a Minkowski vacuum with potentially infinite entropy, the larger theory has an infinite-dimensional Hilbert space. Here, we argue that the QFT analysis applies, and the patch rapidly approaches the vacuum and becomes quiescent, with only a finite number of fluctuations along the way.

This paper is organized as follows:

- In Section 7.2 we define what we mean by "quantum fluctuations," distinguishing between three independent concepts: measurement-induced fluctuations, Boltzmann fluctuations, and vacuum fluctuations. Measurement-induced fluctuations appear when an out-of-equilibrium measuring apparatus interacts with a quantum system, which results in time-dependent branching of the wave function. In contrast, "Boltzmann fluctuations" are inherently dynamical statistical fluctuations, familiar from statistical mechanics. "Vacuum fluctuations," which exist even in stationary states, represent differences between classical and quantum behavior, but do not correspond to dynamical (time-dependent) processes.

- In Section 7.3 we examine eternal de Sitter space in or near the unique Hartle-Hawking vacuum. We first describe the system using QFT in a fixed background. Because the Hartle-Hawking vacuum is stationary, we argue that there are no dynamical fluctuations, despite the fact that an out-of-equilibrium detector (of which there are none present) would measure a nonzero temperature. The cosmic no-hair theorem ensures that all states evolve toward the vacuum, so the system must settle down to a state that is free of dynamical fluctuations. In the context of horizon complementarity, however, each horizon volume can be treated as a system described by a finite-dimensional Hilbert space, and the cosmic no-hair theorem does not apply. If de Sitter space in horizon complementarity is eternal, there will be recurrences and Boltzmann fluctuations, and the conventional picture is recovered.

- In Section 7.4, we turn to models that contain false de Sitter vacua. In semiclassical quantum gravity, or in complementarity in a landscape that includes

a Minkowski vacuum, the dynamics occur in an infinite-dimensional Hilbert space. The situation is then similar to QFT in global de Sitter, where each patch can relax to a stationary quantum state, free of dynamical fluctuations. In complementarity without a Minkowski vacuum, when all vacua are de Sitter, there will still be Boltzmann fluctuations, since the total Hilbert space is finite-dimensional.

- In Section 7.5, we discuss the ramifications of this analysis. First, the conventional Boltzmann brain problem is greatly ameliorated. Even with horizon complementarity, there are no fluctuations in the vacuum to lower-entropy states as long as the larger Hilbert space is infinite dimensional. Similarly, we do not expect uptunneling to higher-energy vacua, which dramatically alters the picture of eternal inflation on a landscape. The standard picture of density fluctuations from inflation remains unchanged, but the understanding of stochastic eternal inflation could be significantly different. Finally, we note that these results depend crucially on one's preferred version of quantum mechanics.

7.2 Fluctuations in Quantum Systems

One way of thinking about the fluctuations of a quantum system is to consider an observable represented by a self-adjoint operator $\widehat{\mathcal{O}}$. If a state $|\Psi\rangle$ is not an eigenstate of $\widehat{\mathcal{O}}$, then the variance

$$(\Delta\widehat{\mathcal{O}})^2_\Psi = \langle\widehat{\mathcal{O}}^2\rangle_\Psi - \langle\widehat{\mathcal{O}}\rangle^2_\Psi \tag{7.2.1}$$

will be strictly positive. Hence, $\widehat{\mathcal{O}}$ does not have a definite value. However, a nonzero variance is not a statement about the *dynamics* of the state, which may well be stationary; it is merely a statement about the distribution of measurement outcomes. In quantum field theory, it is common to refer to radiative corrections from virtual particle pairs as "quantum", "zero-point," or "vacuum" fluctuations, which give rise to phenomena such as the Lamb shift or Casimir effect; they are not, however, "fluctuations" in the sense of a dynamical process that changes the state of the system.

In order to facilitate the investigation of the nature of fluctuations, we define and distinguish between the following types:

Vacuum fluctuations are non-dynamical features of quantum states, which distinguish between classical and quantum behavior, and which ultimately arise as a consequence of the uncertainty principle. In quantum mechanics, vacuum

fluctuations are described by (7.2.1); specifically, in quantum field theory, these fluctuations are understood as radiative corrections from virtual particle pairs.

Measurement-induced fluctuations are fluctuations whose dynamics are generated from a series of measurements of a quantum-mechanical system, resulting in decoherence and wave function branching.

Boltzmann fluctuations are dynamical fluctuations that arise when the microstate of a system is time-dependent, even though the coarse-grained macrostate may be stationary. They are associated, for example, with downward fluctuations in entropy of a thermal macrostate.

To study the Boltzmann brain problem and eternal inflation, we focus on the latter two types of fluctuations, in which time dependence plays a role. For convenience we will use the term "dynamical fluctuations" to refer to either Boltzmann or measurement-induced types. Vacuum fluctuations, of course, are a feature of all quantum systems, but our analysis will not be concerned with such non-dynamical features.

In the remainder of this section we clarify the meanings of "measurement-induced fluctuations," along with the role of measuring devices, and "Boltzmann fluctuations."

7.2.1 Decoherence and Everettian worlds

Let us rehearse the standard understanding of quantum measurement and decoherence in the Everett formulation [7, 171, 172]. There are two underlying postulates of the Everett formulation:

1. The world is represented by quantum states $|\psi\rangle$ that are elements of a Hilbert space $\mathcal{H}$.

2. The time evolution of states is generated by a self-adjoint Hamiltonian operator $\hat{H}$, according to the Schrödinger equation

$$\hat{H}|\psi(t)\rangle = i\partial_t|\psi(t)\rangle \ . \tag{7.2.2}$$

In order to extract a description of a classical world from this formulation, we need to connect observables with measurement outcomes; this connection may be provided by the decoherence program.

As an illustration of decoherence, consider a Hilbert space that factors into an apparatus A that may observe a system S:

$$\mathcal{H} = \mathcal{H}_S \otimes \mathcal{H}_A . \tag{7.2.3}$$

The Schmidt decomposition theorem allows us to write an arbitrary state as

$$|\Psi\rangle = \sum_n c_n |s_n\rangle |a_n\rangle , \tag{7.2.4}$$

where the $|s_n\rangle$ form an orthonormal basis for the system and $|a_n\rangle$ are orthogonal states of the apparatus. We assume that $\dim \mathcal{H}_S < \dim \mathcal{H}_A$, and the sum over n runs up to $\dim \mathcal{H}_S$. The bipartite form of (7.2.4) is unique up to degeneracies in the coefficients $|c_n|$. (For simplicity, we assume there are no degeneracies throughout the remainder of this paper.) Although the Schmidt decomposition identifies a unique basis, there is no mechanism in place to ensure that system and apparatus states are ones that appropriately describe actual measurements. Interactions between the system/apparatus and the environment are crucial for using decoherence to solve the measurement problem.

Incorporating the environment E, the Hilbert space is

$$\mathcal{H} = \mathcal{H}_S \otimes \mathcal{H}_A \otimes \mathcal{H}_E . \tag{7.2.5}$$

It may be possible to write a state in the full Hilbert space using a generalized Schmidt decomposition

$$|\Psi\rangle = \sum_n c_n |s_n\rangle |a_n\rangle |e_n\rangle , \tag{7.2.6}$$

where $|s_n\rangle$ are system basis states; $|a_n\rangle$ are linearly independent, normalized apparatus states; and $|e_n\rangle$ are mutually noncollinear, normalized environment states. The triorthogonal uniqueness theorem [173] guarantees that the form of this tripartite decomposition, if it exists, is unique. (Although this decomposition does not generically exist, it is a necessary feature of the standard decoherence program [7].) Observations are restricted to the system and apparatus, so predictions of the outcomes of measurements are encoded in the reduced density matrix for the system and apparatus, found by tracing out the unobserved degrees of freedom of the environment

from the full density matrix $\rho = |\Psi\rangle\langle\Psi|$:

$$\begin{aligned}
\rho_{SA} &= \text{Tr}_E |\Psi\rangle\langle\Psi| \\
&= \sum_{m,n} c_m c_n^* \langle e_n| e_m\rangle |s_m\rangle|a_m\rangle\langle s_n|\langle a_n| \ .
\end{aligned} \tag{7.2.7}$$

In order for this formalism to describe a quantum state that splits into independent Everettian branches or "worlds," several requirements must be satisfied. First, decoherence must occur—there must be no quantum interference between the different worlds, so observers on one branch evolve independently of the existence of other branches. The absence of interference between states in $\mathcal{H}_S \otimes \mathcal{H}_A$ requires that the reduced density matrix (7.2.7) be diagonal, *i.e.*, that the environment states associated with different branches be orthogonal.

Any density matrix is diagonal in some basis, but that basis might not be a physically viable one, nor one that is in the tripartite form of (7.2.6), where measurement outcomes are accurately reflected in the state of the apparatus. The second requirement is, therefore, that there must exist a basis of apparatus "pointer states" in which decoherence naturally occurs through the dynamical diagonalization of ρ_{SA} in this preferred basis [4, 7, 174–176]. A precise characterization of the pointer states is subtle and context-dependent but roughly corresponds to states of the apparatus that are macroscopically robust (stable). Any interactions between the apparatus and environment should have a minimal effect on the system-apparatus correlations. In principle, we can deduce the pointer states by writing the Hamiltonian as a sum of system/apparatus, environment, and interaction terms:

$$\hat{H} = \hat{H}_{SA} \otimes \mathbb{1}_E + \mathbb{1}_{SA} \otimes \hat{H}_E + \hat{H}_I \ . \tag{7.2.8}$$

The pointer states $|a_n\rangle$ are those whose projectors $\hat{P}_n = |a_n\rangle\langle a_n|$ commute with the interaction Hamiltonian,

$$[\hat{H}_I, \hat{P}_n] = 0 \ . \tag{7.2.9}$$

In practice, the fact that interactions are local in space implies that pointer states for macroscopic objects are those with definite spatial configurations. For instance, if a large object (a billiard ball, a planet, a cat) is in a quantum superposition of two different position eigenstates, interactions with the environment (the air in a room, the cosmic background radiation) will rapidly cause those two possibilities to decohere, creating separate branches of the wave function.

The final feature that is important to describe branching is an arrow of time. We

conventionally imagine that worlds split via decoherence as time passes but almost never merge together, because we implicitly assume that the universe is very far from equilibrium and has evolved from a lower-entropy state in the past. In the present context, "low entropy" means that subsystems begin in a particular state of little or no entanglement, as in (7.2.10). As we demonstrate in the next subsection, dynamical interactions between apparatus and environment naturally increase the amount of entanglement, leading to branching and generating entropy.[2] The standard picture of decoherence and branching is specific to the far-from-equilibrium situation. Near equilibrium, decoherence can arise through rare fluctuations, but is not tied to quantum measurements, as we discuss in Subsection 7.2.3.

7.2.2 Measurement-induced fluctuations

We can use the decoherence program from the previous section to understand the nature of measurement-induced fluctuations. For clarity in the following example, let us identify states in S, A, and E explicitly with subscripts. In the case of real-world quantum measurement, we posit that there is initially no entanglement between any of the factors:

$$|\Psi(t_0)\rangle = |\sigma_*\rangle_S |a_R\rangle_A |e_*\rangle_E \ . \tag{7.2.10}$$

The initial state, denoted by an asterisk, of the system can be arbitrary; but the measuring apparatus must be in a specific "ready" state, denoted by the subscript R. For definiteness, imagine that the system is a single qubit with basis states $\{|+\rangle_S, |-\rangle_S\}$. The apparatus should begin in a ready state and record the results of repeated measurements of the system. We take the apparatus state to be a tensor product of a number of registers (at least one for each measurement we want to perform), where each register is a qutrit with three basis states $\{|+\rangle_A, |-\rangle_A, |0\rangle_A\}$. The ready state of the apparatus is $|a_R\rangle_A = |000\cdots\rangle_A$, and a measurement correlates one of the registers with the state of the system. That is, under unitary evolution we record a measurement in the first register via

$$|+\rangle_S |000\cdots\rangle_A \to |+\rangle_S |+00\cdots\rangle_A \ , \tag{7.2.11}$$

$$|-\rangle_S |000\cdots\rangle_A \to |-\rangle_S |-00\cdots\rangle_A \ . \tag{7.2.12}$$

---

[2]For the purposes of this paper, we are concerned with only the von Neumann entropy from entanglements. There is also the thermodynamic entropy associated with a mixed thermal density matrix, which sets an upper bound on the von Neumann entropy. As the quantum system thermalizes, the von Neumann entropy approaches the thermodynamic entropy [177].

If the apparatus does not start in the ready state, we cannot be confident that it will end up correctly correlated with the state of the system. Since unitary evolution must be reversible, there can be no valid evolution that takes $|+\rangle_S |\psi\rangle_A$ to $|+\rangle_S |+\rangle_A$ for every possible $|\psi\rangle_A$, for example.

Imagine that the system starts in a superposition, so the state takes the form

$$|\Psi(t_0)\rangle = (\alpha|+\rangle_S + \beta|-\rangle_S)\,|000\cdots\rangle_A |e_*\rangle_E \;. \tag{7.2.13}$$

The first step in the evolution is premeasurement, which correlates the apparatus with the system:

$$|\Psi(t_1)\rangle = (\alpha|+\rangle_S |+00\cdots\rangle_A + \beta|-\rangle_S |-00\cdots\rangle_A)\,|e_*\rangle_E \;. \tag{7.2.14}$$

The second step is decoherence, in which the apparatus becomes entangled with the environment:

$$|\Psi(t_2)\rangle = \alpha|+\rangle_S |+00\cdots\rangle_A |e_+\rangle_E + \beta|-\rangle_S |-00\cdots\rangle_A |e_-\rangle_E \;. \tag{7.2.15}$$

Next, we reset in order to perform the measurement again, which means returning the system to its original state. Generally, the environment states will also evolve during this operation. We leave the apparatus unchanged in order to keep a record of the prior measurement outcomes:

$$|\Psi(t_3)\rangle = \alpha|\sigma_*\rangle_S |+00\cdots\rangle_A |\tilde{e}_+\rangle_E + \beta|\sigma_*\rangle_S |-00\cdots\rangle_A |\tilde{e}_-\rangle_E \;. \tag{7.2.16}$$

Finally, we repeat the entire procedure, this time recording the measurement outcome in the second register of the apparatus. After one more iteration of premeasurement and decoherence, we end up with

$$\begin{aligned}
|\Psi(t_4)\rangle = {}& \alpha^2|+\rangle_S |++0\cdots\rangle_A |e_{++}\rangle_E \\
& + \alpha\beta|+\rangle_S |-+0\cdots\rangle_A |e_{-+}\rangle_E \\
& + \alpha\beta|-\rangle_S |+-0\cdots\rangle_A |e_{+-}\rangle_E \\
& + \beta^2|-\rangle_S |--0\cdots\rangle_A |e_{--}\rangle_E \;.
\end{aligned} \tag{7.2.17}$$

At this point the wave function consists of four different decoherent branches, provided that all of the environment states are approximately orthogonal, $\langle e_\mu | e_\nu \rangle_E \approx 0$.

In this context, the statement "we observe quantum fluctuations" is a statement

about measurement-induced fluctuations: it is simply the observation that the history of each individual decoherent branch is one in which the state of the apparatus experiences a time series of observational outcomes, bouncing between $|+\rangle$ and $|-\rangle$. On a randomly chosen branch, the history will exhibit fluctuations between the two outcomes, and all macroscopic objects are robust and physically well-defined (pointer states) by construction. Schrödinger cat superpositions are not allowed, and different worlds or branches must evolve separately.

We see that obtaining the standard measurement outcomes requires both the apparatus to be initially in its ready state and the three Hilbert space factors (system/apparatus/environment) to be initially unentangled. These conditions highlight the crucial role of entropy production in the branching of the wave function and thus in the existence of measurement-induced fluctuations. The reduced density matrix ρ_{SA} has a von Neumann entropy

$$S_{SA} = -\operatorname{Tr}\rho_{SA}\log\rho_{SA} \ . \tag{7.2.18}$$

Since the state as a whole is pure in our example, all of the entropy comes from the entanglement between SA and E. In the initial state (7.2.13), there is no entanglement, and $S_{SA} = 0$. The entropy increases as the state evolves into two branches (7.2.15) and again as it evolves into four branches (7.2.17). Since the entropy of the pure state vanishes, the entropy of the environment equals that of the system/apparatus factor and increases as well. Without entropy production, there are no measurement-induced fluctuations.

Now consider what happens if the entire wave function describing the system, apparatus, and environment (*i.e.*, the whole universe) begins in an energy eigenstate. We assume there are interaction terms in the Hamiltonian that connect the different factors of the Hilbert space. An energy eigenstate obeys

$$\hat{H}|\Psi_n\rangle = E_n|\Psi_n\rangle \ , \tag{7.2.19}$$

where $\hat{H}$ is the full Hamiltonian. Because the wave function is in an energy eigenstate, its time evolution just takes the form of multiplication by an overall time-dependent phase:

$$|\Psi_n(t)\rangle = e^{-iE_n(t-t_0)}|E_n\rangle \ . \tag{7.2.20}$$

The overall phase factor does not affect any of the observable properties of the state; therefore, it is sensible to refer to such a state as "stationary," and its associated

density operator

$$\rho_\Psi = |\Psi_n(t)\rangle\langle\Psi_n(t)| = |E_n\rangle\langle E_n| \tag{7.2.21}$$

is manifestly time independent. Another example of stationary density operator is that of a thermal state with temperature β^{-1}:

$$\rho \sim \exp(-\beta\hat{H}) = \sum_n e^{-\beta E_n}|E_n\rangle\langle E_n| \ . \tag{7.2.22}$$

Indeed, any density matrix diagonal in the energy eigenbasis will be stationary.

In a stationary state, none of the behavior we characterized as "measurement-induced fluctuations"—branching of the wave function into a set of histories with stochastic measurement outcomes over time—is present. In fact, there is no time dependence at all.[3] Certainly, the variance of an observable $\hat{O}$ can be positive in a stationary state, but that variance only leads to dynamical fluctuations if the observable is actually measured. Doing so requires an apparatus that is not itself stationary. Indeed, the apparatus must start in a specific ready state, a condition that we may describe as low entropy. If a quantum state describes the whole universe (as it does in cosmology), and this state is stationary, then it cannot undergo dynamical fluctuations, because nothing can actually change as time passes. For a thermal state in particular, it will be the case that a particle detector beginning in its ready state would detect thermally fluctuating particles; but if all we have to use as a detector is a part of the stationary system itself, it will simply remain stationary, just as the rest of the quantum state does.

7.2.3 Boltzmann fluctuations

There is an important difference between a quantum-mechanical thermal state and one in classical statistical mechanics. Classically, a state in thermal equilibrium has a uniform temperature in space that is also constant in time. However, this description is macroscopic and obtained by coarse graining. Any realization of such a system

---

[3]Even in stationary states, one can define an effective evolution with respect to correlations with a clock subsystem [178]. The effective time parameter τ has nothing to do with the ordinary coordinate time t; all such time evolutions are present at every moment of (ordinary) time. From this perspective, a large number of Boltzmann brains and similar fluctuations actually exist at every moment in an apparently stationary spacetime. Such a conclusion would apply to Minkowski spacetime as well as to de Sitter, in conflict with the conventional understanding that dynamical fluctuations in de Sitter depend on the Gibbons-Hawking temperature (but see [179, 180]). This kind of effective evolution is fundamentally different from the ordinary evolution studied in this paper.

with nonzero temperature has a microstate that is time-dependent. For instance, the atoms and molecules in a box of gas are individually moving, even if the temperature and density are constant. The system will, therefore, undergo rare fluctuations to nonequilibrium states. The probability of observing such a fluctuation to a state with entropy ΔS lower than equilibrium scales as $\sim e^{-\Delta S}$. To avoid confusion we refer to such events, in which the evolution of the microstate causes a reduction in entropy, as "Boltzmann fluctuations," to distinguish them from "measurement-induced fluctuations" where the wave function branches, which increase von Neumann entropy.

In quantum mechanics, individual energy eigenstates are stationary, in contrast with classical states of nonzero energy. Stationary quantum states will not experience Boltzmann fluctuations. A statistical ensemble of stationary states will itself be stationary; we expect no Boltzmann fluctuations there as well. In particular, a closed system in a mixed thermal state, with a stationary density operator $\rho \sim e^{-\beta \hat{H}}$, should not have Boltzmann fluctuations when regarded as a statistical ensemble of energy eigenstates. However, we most commonly encounter thermal density matrices after tracing over environmental degrees of freedom. In that case the remaining system is not closed, and we need to be a bit more careful.

Consider a decomposition of a closed quantum system into a set of macroscopically observable system variables and an environment:

$$\mathcal{H} = \mathcal{H}_S \otimes \mathcal{H}_E \ . \tag{7.2.23}$$

(We have absorbed the apparatus that appears in (7.2.5) into our definition of the macroscopic system.) The environment includes local but microscopic variables (such as the positions and momenta of individual gas molecules, as opposed to macroscopic fluid variables such as temperature and pressure), as well as causally disconnected degrees of freedom (such as modes outside a cosmological horizon). Expectation values of macroscopic observables in a pure state $|\Psi\rangle \in \mathcal{H}$ are encoded in the reduced density matrix $\rho_S = \text{Tr}_E |\Psi\rangle\langle\Psi|$, with entropy given by $S_S = -\text{Tr}\,\rho_S \log \rho_S$. While the evolution of the pure state $|\Psi\rangle$ is unitary, that of ρ_S is generally not. It is described by a Lindblad equation [181], which allows for transfer of information between the macroscopic system and the environment:

$$\dot{\rho}_S = i[\hat{H}_*, \rho_S] + \sum_n \left(\hat{L}_n \rho_S \hat{L}_n^\dagger - \frac{1}{2}\hat{L}_n^\dagger \hat{L}_n \rho_S - \frac{1}{2}\rho_S \hat{L}_n^\dagger \hat{L}_n \right) \ . \tag{7.2.24}$$

The Lindblad operators $\hat{L}_n$ characterize the non-unitary part of the evolution of the

$$\begin{pmatrix} x & & & & & \\ & 0 & & & & \\ & & 0 & & & \\ & & & 0 & & \\ & & & & 0 & \\ & & & & & 0 \end{pmatrix} \longrightarrow \begin{pmatrix} x & & & & & \\ & x & & & & \\ & & x & & & \\ & & & x & & \\ & & & & 0 & \\ & & & & & 0 \end{pmatrix} \longrightarrow \begin{pmatrix} \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix} \longrightarrow \left(\begin{array}{c|c} x & 0 \\ \hline & \begin{matrix} \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{matrix} \end{array} \right)$$

Figure 7.1. Schematic evolution of a reduced density matrix in the pointer basis. The density matrix on the left represents a low-entropy situation, where only a few states are represented in the wave function. There are no off-diagonal terms, since the pointer states rapidly decohere. The second matrix represents the situation after the wave function has branched a few times. In the third matrix, the system has reached equilibrium; the density matrix would be diagonal in an energy eigenbasis, but in the pointer basis, decoherence has disappeared and the off-diagonal terms are nonzero. The last matrix represents a Boltzmann fluctuation in which one pointer state has fluctuated into existence by decohering with respect to the other states.

system as induced by interactions with the environment, and will depend on the specific setup being studied. The Hermitian operator $\hat{H}_*$ is not necessarily equal to the self-interaction Hamiltonian of the system alone; it captures the part of the entire Hamiltonian that induces unitary evolution on the reduced density matrix, including possible renormalization effects due to interaction with the environment. A system far from equilibrium will generally exhibit dissipation and entropy increase (see *e.g.* [182]), and we may define a dissipation timescale on which the system will approach a stationary state.

On much longer timescales, however, even systems with approximately-stationary reduced density matrices can experience decoherence and transitions to lower-entropy states, in precise analogy with Boltzmann fluctuations in classical statistical mechanics. In Figure 7.1 we provide a schematic representation of the evolution of the reduced density matrix, written in the pointer basis. The first two entries show the splitting of branches of the wave function starting from a low-entropy configuration, as described for example by the transition from (7.2.13) to (7.2.15) and to (7.2.17). The state branches and decoheres, remaining diagonal in the pointer basis. Eventually, it approaches equilibrium and becomes diagonal in the energy eigenbasis; by that point, the off-diagonal elements in the pointer basis are comparable to the diagonal ones, and the pointer states are no longer decoherent. From equilibrium, there can be rare fluctuations (if the total Hilbert space is finite-dimensional) to lower-

entropy configurations where one branch has once again decohered from the rest, as shown in the last entry.

Crucially, the existence of such fluctuations depends on the dimensionality d_E of the Hilbert space $\mathcal{H}_E$ of the environment (assumed to be larger than the dimensionality of the system's Hilbert space $\mathcal{H}_S$). For finite d_E, Hilbert space is bounded, and one can derive a quantum version of the Poincaré recurrence theorem [164]; for infinite d_E, the recurrence time goes to infinity, and excitations in the system can dissipate into the environment and never come back. Zurek [183] has shown that, under reasonable assumptions concerning the initial wave function and the distribution of eigenvalues, the correlation amplitudes governing off-diagonal elements in the reduced density matrix will have an average of zero and experience fluctuations with a magnitude that scales as

$$\Delta \sim d_E^{-1/2} \ . \tag{7.2.25}$$

In a finite-dimensional Hilbert space, Boltzmann fluctuations are inevitable; however, in an infinite-dimensional space, the system can settle into equilibrium and stay there forever. The reduced density matrix corresponding to the latter asymptotes to a stationary form, free of Boltzmann fluctuations.

This discussion presumes that the branching structure of the wave function can be discerned from the form of the reduced density matrix for the macroscopic variables $\mathcal{H}_S$. In general, we cannot tell what states of a quantum system are actually realized on different branches simply by looking at its reduced density matrix.[4] For example, we might have a single qubit that takes on different states on three different branches of the wave function, specified by three mutually orthogonal environment states:

$$|\Psi\rangle = \frac{1}{\sqrt{2}}|+z\rangle_S|e_\uparrow\rangle_E + \frac{1}{2}|+x\rangle_S|e_\rightarrow\rangle_E - \frac{1}{2}|-x\rangle_S|e_\leftarrow\rangle_E \ . \tag{7.2.26}$$

The reduced density matrix for the qubit is

$$\rho_S = \frac{1}{2}|+z\rangle_S\langle+z| + \frac{1}{4}|+x\rangle_S\langle+x| + \frac{1}{4}|-x\rangle_S\langle-x| \tag{7.2.27}$$

$$= \frac{3}{4}|+z\rangle_S\langle+z| + \frac{1}{4}|-z\rangle_S\langle-z| \ . \tag{7.2.28}$$

In the last line, the existence of three branches is completely obscured; the reduced density matrix does not reveal which states of the system exist as part of distinct

---

[4]We thank Alan Guth, Charles Bennett, and Jess Riedel for discussions on this point.

worlds.

Thus, the reduced density matrix alone is not enough information to reveal what is truly happening inside a system. Indeed, it is possible to construct a stationary reduced density matrix from an appropriate mixture of nonstationary states by tracing out the environment. Therefore, the fact that a reduced density matrix is stationary does not suffice to conclude that there are no dynamical processes occurring on distinct branches within the system that it describes; for that, it is necessary to consider the full quantum state. When we discuss the thermal nature of a patch of de Sitter space in Section 7.3.1, we have the benefit of knowing the full state of the de Sitter vacuum, allowing us to circumvent this issue and draw conclusions about the (lack of) dynamics in a patch.

7.3   Single de Sitter Vacua

We now apply these ideas to de Sitter cosmology—specifically, to the case of a unique vacuum with $\Lambda > 0$. In the Hartle-Hawking vacuum, the quantum state of any one causal patch is described by a thermal reduced density matrix. As emphasized in Subsection 7.2.3 above, we cannot claim that the patch is stationary on the sole basis of its reduced density matrix; however, given that we know the full vacuum state, we argue that the patch is indeed stationary. Were we to observe the patch, we would see fluctuations, but in the absence of an external observing device, nothing fluctuates. In particular, there are no decohered branches of the wave function containing time-series records of fluctuating observables. This picture does not apply if horizon complementarity is valid; in this case the entire Hilbert space is finite-dimensional, and unless it starts there, the state cannot asymptote to the vacuum as $t \to \infty$. In complementarity, we expect Boltzmann fluctuations and Poincaré recurrences.

7.3.1   Eternal de Sitter

Let us recall some basic properties of quantum fields in de Sitter space [168, 184]. De Sitter space is the unique maximally symmetric spacetime with positive curvature. In 4D, it has a scalar curvature $12H^2$ and satisfies the Einstein equations with a cosmological constant $\Lambda = 3H^2$, where H^{-1} is the radius of de Sitter space. Consider

a massive[5], noninteracting scalar field φ, which satisfies the Klein-Gordon equation

$$(\Box - m^2)\varphi = 0 \qquad (7.3.1)$$

in the de Sitter metric. In order to quantize fields in de Sitter space, we must first choose a coordinate system. There are numerous possibilities, but we narrow the scope to flat coordinates and static coordinates, as they are used most often in the literature.

In flat coordinates, the metric reads

$$ds^2 = \frac{1}{H^2\tau^2}\left(-d\tau^2 + dx_i dx^i\right) , \qquad (7.3.2)$$

which has the form of a flat, expanding Friedmann-Robertson-Walker metric with a constant Hubble parameter H and conformal time τ. In these coordinates, there is no timelike Killing vector to provide a sensible prescription for defining modes of φ. Since there is translational and rotational invariance among the spatial directions, we are still able to separate the mode solutions with wave number $\vec{k}$ as

$$f(\tau)e^{i\vec{k}\cdot\vec{x}} \qquad (7.3.3)$$

for some function f. Thus, we may attempt to define modes in the asymptotic regions of de Sitter, $\mathcal{I}^{\pm}$, by analogy with Minkowski space. Because of this analogy, the vacuum defined by these modes will have the same symmetries as the free field Minkowski vacuum. Unfortunately, the asymptotic regions are not static in an expanding universe, so we are left to define modes in the adiabatic approximation for a universe that has an infinitely slow expansion. The Euclidean vacuum, formed from the adiabatic modes, is invariant under the de Sitter group and, thus, does not change with time. Although de Sitter invariance alone does not define a unique state, the Euclidean vacuum is the unique de Sitter-invariant Hadamard[6] state for a massive, noninteracting scalar field [185, 188–192].

---

[5]We do not consider the massless case, since there is no (vacuum) state that is invariant under the full de Sitter group [185], which is problematic for the cosmic no-hair theorem in Subsection 7.3.2. However, if one assumes the shift invariance of the massless scalar field is just a global gauge transformation, then a fully de Sitter invariant vacuum can in fact be defined [186].

[6]Relaxing the Hadamard condition [187] yields a continuous family of de Sitter-invariant states, known as the α vacua, which are related to one another via Bogoliubov transformations [185].

In static coordinates the metric becomes

$$ds^2 = -\left(1 - H^2 r^2\right) dt^2 + \left(1 - H^2 r^2\right)^{-1} dr^2 + r^2\, d\Omega^2 \ . \qquad (7.3.4)$$

These coordinates give a timelike Killing vector $-\partial_t$ that points toward the future (past) in the northern (southern) causal diamond, and we may use this Killing vector to define modes. Following [193], the mode expansions for the southern and northern diamonds of de Sitter space are

$$\varphi^S = \int_0^\infty d\omega \sum_{j=-\infty}^\infty \left[a_{\omega j}^S \varphi_{\omega j}^S + \left(a_{\omega j}^S\right)^\dagger \left(\varphi_{\omega j}^S\right)^* \right] \qquad (7.3.5)$$

$$\varphi^N = \int_0^\infty d\omega \sum_{j=-\infty}^\infty \left[a_{\omega j}^N \varphi_{\omega j}^N + \left(a_{\omega j}^N\right)^\dagger \left(\varphi_{\omega j}^N\right)^* \right] \ , \qquad (7.3.6)$$

where ω is the mode frequency. The operators $a_{\omega j}^N$ and $\left(a_{\omega j}^S\right)^\dagger$ are annihilation operators in the northern and southern diamonds. The Euclidean vacuum is

$$|\Omega\rangle = \prod_{\omega=0}^\infty \prod_{j=-\infty}^\infty \left(1 - e^{-2\pi\omega}\right)^{1/2} \exp\left[e^{-\pi\omega} \left(a_{\omega j}^N\right)^\dagger a_{\omega j}^S \right] |S\rangle \otimes |N\rangle \ , \qquad (7.3.7)$$

where $|S\rangle$ and $|N\rangle$ are the southern and northern no-particle vacua. Ignoring gravitational back-reaction, the static Hamiltonian associated with the northern modes is

$$\hat{H}_N = \int_0^\infty d\omega \sum_{j=-\infty}^\infty \left(a_{\omega j}^N\right)^\dagger a_{\omega j}^N\ \omega \ , \qquad (7.3.8)$$

and the reduced density matrix in the northern diamond is

$$\rho_N = \mathrm{Tr}_S\, |\Omega\rangle\langle\Omega| = \left[\prod_\omega (1 - e^{-2\pi\omega}) \right] e^{-\beta \hat{H}_N} \ , \qquad (7.3.9)$$

which is a thermal density matrix with temperature $T = 1/\beta$.

If the universe is in the Euclidean vacuum, the reduced density matrix describing the area inside a causal horizon is thermal. In Subsection 7.2.3, we argued that a subsystem with a thermal density matrix may still evolve into one with a Boltzmann fluctuation. In the case of the Euclidean vacuum, however, we have both the reduced density matrix ρ_N and the full quantum state $|\Omega\rangle$. From an examination of (7.3.7), we see that the modes of a given frequency ω in the northern diamond are in a

one-to-one correspondence with the modes in the southern diamond. By tracing out the southern diamond to construct ρ_N, we know precisely which correlations we are discarding, mode by mode. Furthermore, there is no interaction Hamiltonian between the northern and southern diamonds, since the diamonds are not in causal contact. The entanglement structure is not disrupted by the separate evolution in each diamond, so dynamical processes, akin to the one shown in the last panel of Figure 7.1, are forbidden. Then the reduced density matrix of each diamond is truly stationary, and no Boltzmann fluctuations are possible in either diamond. (The spacetime geometry does not necessarily approach de Sitter globally, but it asymptotes toward stationarity in each patch, which is all we really need.)

We have argued that there are no Boltzmann fluctuations in the de Sitter vacuum. It remains to determine whether the universe may actually be described by the de Sitter vacuum. Accordingly, the rest of our analysis consists of understanding the conditions under which the quantum state takes on this stationary vacuum form in different models.

7.3.2 Cosmic no-hair

We turn now to situations, like that of our universe today, in which the universe is not in the vacuum but rather evolving in time. We will see that, though there may be dynamical fluctuations initially if the state is very far from the vacuum, the state of a single patch will quickly approach the vacuum on time scales proportional to the inverse of the Hubble parameter, after which no fluctuations will arise.

We begin with the classical form of the cosmic no-hair theorem, which states that, given a positive vacuum energy density (*i.e.*, a positive cosmological constant Λ), the metric evolves locally toward that of de Sitter space [8]. Physically, excitations of de Sitter (including matter and radiation fields with substantial energy densities) redshift away across the horizon, so every causal patch relaxes to the vacuum.

The physical intuition behind the cosmic no-hair theorem extends to quantum fields in curved spacetime. For generic states, the expectation value of a massive scalar field φ decays exponentially in time:

$$\langle \varphi(x) \rangle_\psi = \mathcal{O}(e^{-M|\tau|}) \,, \tag{7.3.10}$$

for a decay constant $M > 0$ and proper time τ between the point x and some reference point at $\tau \to \infty$ [9]. Higher n-point correlation functions at large separations decay as well. The vacuum is stable against perturbations and is an attractor state for local

operators, whose expectation values in a generic state will approach the expectation values in the vacuum in the asymptotic region [194].

A quantum-gravitational version of the no-hair theorem would presumably yield analogous results for the graviton field $h_{\mu\nu}$, but a scalar field can stand in as a proxy in order to make calculations manageable. Although we have focused on a free scalar field theory to write an explicit form of the Euclidean vacuum and the reduced density matrix, the graviton has self interactions, so the analysis needs to be extended to an interacting scalar theory with a Hartle-Hawking vacuum. For renormalizable interactions, the cosmic no-hair theorem still holds at an arbitrary number of loops, for arbitrary n-point functions, and for $D \geq 2$. Furthermore, M does not receive any radiative corrections. The results of [9, 10] show that the decay constant for massive[7] scalar fields is

$$M = \begin{cases} \frac{3}{2}H & \text{for } m > \frac{3}{2}H \\ \frac{3}{2}H - \sqrt{\frac{9}{4}H^2 - m^2} & \text{for } 0 < m \leq \frac{3}{2}H \end{cases} \tag{7.3.11}$$

If the universe is in an arbitrary state that is perturbed around the Hartle-Hawking vacuum, the state will approach the vacuum at large spacetime distances exponentially fast, with a decay constant $3H/2$ for large m. Once the field correlations have sufficiently decayed, the arguments of Subsection 7.2.3 tell us that no dynamical fluctuations occur.

7.3.3   Complementarity in eternal de Sitter

Horizon complementarity posits that the spacetime interpretation of a quantum state depends on the viewpoint of a specified observer [150, 169, 170, 263]. In particular, a description in terms of local quantum field theory will not extend smoothly beyond a horizon. Applied to de Sitter space, this philosophy implies that spacetime locality only applies within a cosmological horizon volume, and the corresponding quantum system has a finite-dimensional Hilbert space. The Hilbert space of the patch can be decomposed as a product of bulk and boundary factors [198, 279]:

$$\mathcal{H} = \mathcal{H}_{\text{bulk}} \otimes \mathcal{H}_{\text{boundary}} . \tag{7.3.12}$$

---

[7]As previously mentioned, the massless case is problematic, since there is no de Sitter-invariant vacuum in the noninteracting limit [185]. With nonvanishing interactions, correlation functions of the field at large timelike separations grow no faster than a polynomial function of $H\tau$ at the perturbative level [195]. There is, however, evidence at one and two loops that the 2-point correlation function decays as a polynomial of H [195–197].

(We ignore a possible factor corresponding to singular spacetime geometries, which will not be important for our analysis.)

From the Bekenstein-Hawking relation [199, 200], the entropy associated with the patch is one quarter of the area of the horizon: $S_{\text{dS}} = \mathcal{A}/4$. This entropy is related to the density matrix $\rho \sim e^{-\beta\hat{H}}$ for the patch via $S_{\text{dS}} = -\operatorname{Tr}_{\text{boundary}} \rho \ln \rho$, so the patch is thermal even if the system as a whole is in a pure quantum state. The energy spectrum is discrete, with only a finite number of eigenvalues with energies less than any given cutoff value [201].

If we interpret the entropy as being the logarithm of the number of quantum states, the horizon patch is analogous to a closed thermal system at a temperature T [149, 150]. Although the relationship $\dim\mathcal{H} = e^S$ holds only at infinite temperature, there are compelling reasons (*e.g.*, from black holes) to think that the static Hamiltonian is bounded from above [202].[8] In our discussion of complementarity, we assume that this bound exists and that the dimension of the Hilbert space

$$\dim\mathcal{H} = e^{2S_{\text{dS}}} = \exp(6\pi\Lambda^{-1}) \tag{7.3.13}$$

is finite. (The factor of 2 comes from the fact that the bulk and boundary components have equal dimensionality.)

The complementarity picture of eternal de Sitter with a unique vacuum state is, therefore, very different from the situation of QFT in a de Sitter background discussed in Subsection 7.3.1. In the latter, the ability of excitations to leave the horizon and never return depended crucially on the fact that Hilbert space was infinite-dimensional. In complementarity, eternal de Sitter space is a truly closed finite-dimensional system, subject to Poincaré recurrences [164]. Of course, there is a true vacuum state, the lowest-energy eigenstate, that is strictly stationary, but a generic state is nonstationary. We may think of excitations as being absorbed by a stretched horizon with a finite area and eventually being emitted back into the bulk, as shown in Figure 7.2. Boltzmann fluctuations into lower-entropy states (described in Subsection 7.2.3) are allowed, in agreement with the conventional picture of a thermal de Sitter patch. As we argue below, this story changes in important ways in theories with more than one metastable vacuum.

---

[8]For subtleties involving the use of the static Hamiltonian in quantum gravity, see [203].
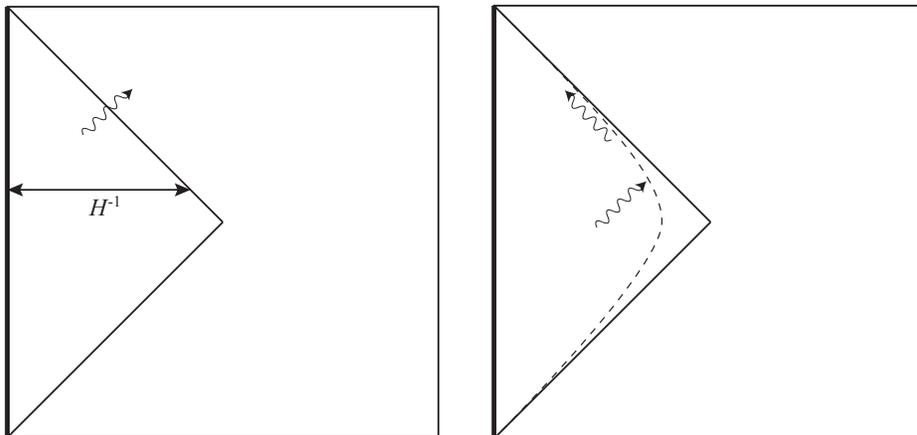
Figure 7.2. Conformal diagrams for de Sitter space in the global (QFT) picture [left] and with horizon complementarity [right]. We consider an observer at the north pole, represented by the line on the left boundary and their causal diamond (solid triangle). The wavy line represents excitations of the vacuum approaching the horizon. In QFT in curved spacetime, the excitation exits and the state inside the diamond approaches the Hartle-Hawking vacuum, in accordance with the cosmic no-hair theorems. In contrast, horizon complementarity implies that excitations are effectively absorbed at the stretched horizon (dashed curve just inside the true horizon) and eventually return to the bulk.

7.4   Multiple Vacua

In this section we consider theories with more than one metastable potential minimum, at least one of which has $\Lambda > 0$, as portrayed schematically in Figure 7.3. We consider the existence of dynamical fluctuations in both the lowest-energy "true" vacuum and in any higher-energy false vacua. For convenience, we limit our attention to vacua with non-negative energy, $\Lambda \geq 0$. Transitions from vacua with $\Lambda \geq 0$ to those with $\Lambda < 0$ generally result in singular crunches; evolution might continue via quantum-gravity effects, but we will not address that possibility here.

7.4.1   Semiclassical quantum gravity

We first consider semiclassical quantum gravity, by which we mean QFT coupled to a classical (but dynamical) spacetime background. Coleman studied false vacua in this context and calculated the rate at which a higher-energy vacuum would decay to a lower-energy state via bubble nucleation [204, 205]. It is useful to consider an
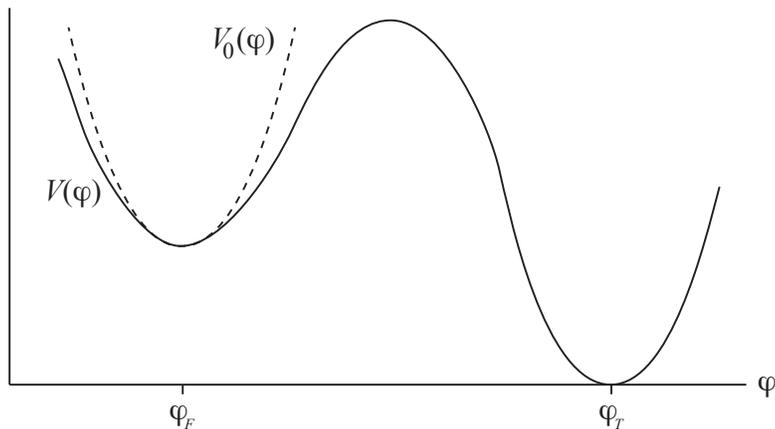
Figure 7.3.   A scalar field potential with multiple local minima. The global minimum corresponds to the true-vacuum value φ_T (which may have $\Lambda = 0$ or $\Lambda > 0$), and for simplicity we have portrayed a single false-vacuum value φ_F. The dashed line represents the perturbative Hamiltonian for the false vacuum, in which the potential is given by a local approximation to the true potential in the vicinity of φ_F.

analogous problem in one-dimensional quantum mechanics, in which a single particle moves in a potential $V(x)$, with a global (true) minimum at x_T and a local (false) minimum at x_F. Then, one can calculate the transition amplitude using the path integral defined with respect to Euclidean time T:

$$\left\langle x_T \middle| e^{-HT} \middle| x_F \right\rangle = N \int [dx] e^{-S_E[x(T)]} \ , \tag{7.4.1}$$

where H is the Hamiltonian and S_E is the Euclidean action, while the states $|x_T\rangle$ and $|x_F\rangle$ are position eigenstates. This quantity can be calculated using instanton methods and represents the amplitude for finding the particle at position x_T, given that it started at position x_F—something that might be of relevance to an observer measuring the position of the particle. An analogous field theory calculation can be used to calculate the rate of transition from one field configuration $|\varphi_1(x)\rangle$ to another $|\varphi_2(x)\rangle$, including the tunneling rate from one vacuum to another, as shown in Figure 7.3.

   Our interest, however, is not in what an out-of-equilibrium observer with a field-value detection device would measure, but in how quantum states evolve in isolated patches of de Sitter space. Eigenstates of the field operator $\hat{\varphi}(x)$ are not energy

eigenstates; therefore, we need to be careful when we use terms such as "false vacuum" and "true vacuum" to refer to quantum states rather than field values. For some purposes it is useful to study eigenstates of a perturbative Hamiltonian constructed by approximating the potential in the vicinity of one local minimum, as shown for φ_F in Figure 7.3. In that case the results from Section 7.3.1, where we studied QFT in a fixed de Sitter background, are relevant.

Consider first the true vacuum quantum state $|0\rangle$ of the full theory. A generic homogeneous field value φ_* will have some nonzero overlap with this state, $\langle \varphi_* | 0 \rangle \neq 0$, but the field will be mostly localized near the global minimum value φ_T. While it is difficult to rigorously prove a version of the cosmic no-hair theorem for this interacting theory, we intuitively expect the physics in this case to mirror that of QFT with a unique de Sitter vacuum. Namely, excitations above the lowest-energy state will dissipate outside the horizon, and each local patch will approach the vacuum state $|0\rangle$. This state is stationary, and we expect no measurement-induced or Boltzmann fluctuations. Since we are dealing with QFT, the Hilbert space is infinite-dimensional, and there are no recurrences.

We also do not expect uptunneling to a higher-energy vacuum from the true vacuum state for the same reason (energy eigenstates are stationary and do not fluctuate). This assertion might seem to be in tension with the existence of instantons that contribute a nonzero amplitude to processes analogous to (7.4.1), but such a counterargument confuses field values with quantum states. Although there are instanton solutions, their role is to shift the value of the vacuum energy in the true vacuum from what one would compute in a local approximation to the effective potential near φ_T. As noted above, the nonzero overlap between two perturbative vacua can be interpreted as a transition rate between them. But we are interested in the states of definite semiclassical geometry, which should correspond to vacua of the full potential, where instanton corrections have already been taken into account.

The situation is analogous to that of the QCD vacuum, where instantons connecting vacua of different winding numbers provide a shift in energy that depends on the value of θ_{QCD}. The QCD vacuum is a single, static state which incorporates the instanton corrections, not constantly occurring dynamical transitions between states of definite winding number, just like a harmonic oscillator in an energy eigenstate is static rather than undergoing constant fluctuations. Even though we can write the QCD vacuum in a basis of states of different winding number, or an energy eigenstate of the harmonic oscillator in a basis of position eigenstates, the lesson of Section 7.2 is that such descriptions have no physical reality. Instantons are important for cal-

culating energy eigenvalues, but once the quantum system is in a stationary state such as the vacuum $|0\rangle$, they do not describe true dynamical transitions. The local perturbative vacuum will be unstable to uptunneling via instantons, but that's not the true nonperturbative vacuum into which the system settles.

Next we turn to false vacua. A semiclassical state with $\langle\varphi\rangle = \varphi_F$ is not strictly a vacuum state, or indeed any form of energy eigenstate, as it will decay via tunneling. We may nevertheless consider the energy eigenstates of the perturbative Hamiltonian, obtained by locally approximating the potential in the vicinity of φ_F, as shown in Figure 7.3. These are not energy eigenstates of the full Hamiltonian, but their dynamics are well-described by a combination of processes near the false-vacuum value plus decays via bubble nucleation. We may think of the "false de Sitter vacuum state" as the Hartle-Hawking vacuum state of this perturbative Hamiltonian. Once again, we expect excitations to rapidly dissipate by leaving the horizon, resulting in a state that does not exhibit thermal fluctuations. We refer to such states as "quiescent" (reserving the term "stationary" for true energy eigenstates).

We are left with two kinds of possible non-perturbative processes to consider: downtunneling to lower-energy vacua and uptunneling to higher-energy vacua. First, we examine downtunneling. In the conventional picture of false-vacuum decay, a small bubble of true vacuum nucleates and grows at nearly the speed of light. This picture is clearly a semiclassical description of a single branch of the wave function, rather than a full treatment of the quantum state. We can decompose the Hilbert space into the product of the state space of a smooth background field $\varphi_\lambda(x)$ and small-scale fluctuations:

$$\mathcal{H} = \mathcal{H}_{\varphi_\lambda} \otimes \mathcal{H}_{\delta\varphi} . \tag{7.4.2}$$

Here, λ is a length scale used to smooth the field. The factor $\mathcal{H}_{\varphi_\lambda}$ includes configurations with bubbles of different sizes and locations, as well as completely homogeneous configurations. When a bubble nucleates, some of the energy density that was in the potential for φ gets converted into fluctuation modes, resulting in the production of entropy. Therefore, a reduced density matrix for the background field obtained by tracing over $\mathcal{H}_{\delta\varphi}$ will exhibit decoherence, as the fluctuations produced by bubbles in different locations will generically be orthogonal to each other. In that sense, the semiclassical configurations described by bubble nucleation correspond to truly distinct branches of the wave function. With that single caveat, we agree with the standard picture of downtunneling to lower-energy vacua.

Different cases of interest for bubble nucleation are shown in Figure 7.4. An observer at the north pole in the de Sitter diagram could witness the nucleation of a
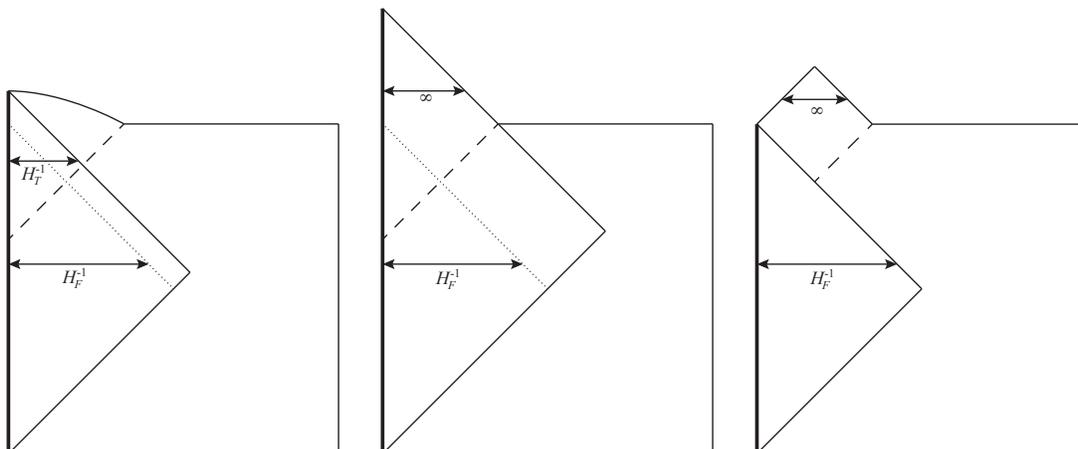
Figure 7.4.   Conformal diagrams for de Sitter space with a false vacuum. The first two diagrams show the effect of a bubble (dashed line) nucleating within the northern-hemisphere causal patch, leading to lower-energy de Sitter [left] and Minkowski [middle], while the third shows a false-vacuum region that does not experience any bubbles [right]. H_F^{-1} and H_T^{-1} are the Hubble radii of the false and true vacua, with the latter being infinite in the Minkowski case. In both cases the true horizon is larger than the Hubble radius in the false vacuum; in the left-hand diagram, it becomes equal to the horizon in the true vacuum, while on the right it becomes infinitely large. In either of these cases, excitations can leave the apparent horizon in the false vacuum while remaining inside the true horizon. On the right, the observer at the North pole remains in the false vacuum state forever, although there are bubbles outside their horizon.

bubble to a lower-energy de Sitter vacuum, or to a Minkowski vacuum (the triangular "hat"), or avoid seeing bubbles at all. The probability of seeing a bubble along any specified geodesic asymptotes to 1, but for a sufficiently small nucleation rate, the physical volume of space remaining in the false vacuum grows with time.

Next we turn to uptunneling from one false-vacuum state to another of even higher energy.[9] In the true vacuum, we could straightforwardly argue that the spirit of the cosmic no-hair theorem is obeyed: excitations leave the horizon and the system approaches its lowest-energy eigenstate. In the false vacuum, the argument is not so clean, since there are no true energy eigenstates to approach. Nevertheless, the physical situation is quite similar. The Hilbert space is still infinite-dimensional, so we do not expect recurrences, and excitations within a patch can readily leave the horizon,

---

[9]We thank Stefan Leichenauer and Paul Steinhardt for discussions of these issues.

leaving us in the perturbative vacuum. Once again, there exist instantons reflecting the overlap of the perturbative false vacuum with excited states. Just as above, however, we cannot interpret these instantons as true dynamical processes; they signal, instead, that the actual state of the field in the false vacuum is not entirely concentrated around φ_F, but also has some support on field values corresponding to potential minima of higher energies. We said above that excitations around the perturbative false vacuum state dissipate away, resulting in a state quiescent with respect to perturbative fluctuations. Incorporation of instanton effects simply shifts the quiescent state slightly, like it did for the true vacuum state. It is this quiescent state that is physical, not the perturbative vacua connected by instanton transitions.

We can see, in particular, that the wave function cannot start in a quiescent state and then split into two branches, describing uptunneling: uptunneling represents a decrease rather than an increase of entropy, so it can only be a Boltzmann fluctuation rather than a branching of the wave function. The timescale over which the perturbative vacuum relaxes to the physical one (presumably with slightly smaller effective cosmological constant) will be governed by the barrier-penetration factor connecting the false vacuum to higher-energy minima. That factor also governs the rate for uptunneling to such minima. Therefore, we expect a relatively short window in which uptunneling can happen before the state relaxes, after which the rate of uptunneling falls to zero. We can check this picture by considering the limit in which the barrier between false and true vacuum becomes infinitely large. In that case, transitions from the false to true vacua are suppressed, and the behavior of the false vacuum should increasingly resemble our picture of the true vacuum. This is precisely what we have found: in both cases, uptunneling is forbidden (more precisely, the rate of uptunneling is suppressed as excitations dissipate away).

While these results are not rigorous, they provide a strong indication that false-vacuum states in semiclassical quantum gravity either decay or asymptote to quiescent states that are free of dynamical fluctuations.

7.4.2 Complementarity in a landscape

We now consider theories with multiple vacua, each labeled by a field expectation value φ_i, in the context of horizon complementarity. In this case the Hilbert space appropriate to a single vacuum (7.3.12) is promoted to a direct sum, with one term for each semiclassical patch geometry:

$$\mathcal{H} = \bigoplus_i \mathcal{H}^{(i)}_{\text{bulk}} \otimes \mathcal{H}^{(i)}_{\text{boundary}} \ . \tag{7.4.3}$$

The structure is similar to that of Fock space [198, 279]. The dimensionality of the entire Hilbert space is the sum of the dimensions of each term, $\dim \mathcal{H}^{(i)} = e^{2S_{\mathrm{dS}}^{(i)}} = \exp(6\pi\Lambda_i^{-1})$. There are two cases of interest: the finite-dimensional case where every vacuum has $\Lambda > 0$ and the infinite-dimensional case where there is at least one vacuum with $\Lambda = 0$. (As mentioned previously, we do not consider vacua with $\Lambda < 0$, as transitions into them lead to singularities.)

If all vacua have $\Lambda > 0$, the situation is very similar to the single-vacuum case discussed in Subsection 7.3.3. Exact energy eigenstates, including the lowest-energy vacuum state, will be stationary and no dynamical fluctuations will occur. The vacuum will feature a de Sitter semiclassical geometry with the field value concentrated near the true minimum, although it will not be a field eigenstate. Generic states, however, will not be stationary, and in a finite-dimensional Hilbert space there is no room for excitations to dissipate outside the horizon, so recurrences are expected.

Now consider theories with at least one vacuum having $\Lambda = 0$, as might be expected in supersymmetric or string theories. The future development of the spacetime includes census-taker observers living in a Minkowski "hat" [206, 207], as shown in the middle and right diagrams of Figure 7.4. The Hilbert space of the full theory is then infinite dimensional, and such observers have access, in principle, to an infinite amount of information.

From (7.2.25), the rate of Boltzmann fluctuations goes to zero (and the timescale for recurrences goes to infinity) for infinite-dimensional Hilbert spaces, where $\Lambda_T = 0$. Of course, there are no dynamical fluctuations in the true Minkowski vacuum. But we can make a stronger statement: the rate of the fluctuations will asymptote to zero even in the false vacua. The intuition is that states with excitations around false-vacuum geometries are more likely to decay than the vacuum states themselves. So time evolution will skew the population of false vacua towards states that are stationary except for the possibility of decay by bubble nucleation, *i.e.* quiescent in the sense of the previous subsection. After a high-energy vacuum decays to a lower-energy one, transient excitations will allow for the existence of Boltzmann fluctuations, but the excited states will again preferentially decay. The surviving configurations will become effectively stationary, and the Boltzmann fluctuation rate will asymptote to zero, rather than to a nonzero constant. We therefore expect only a finite (and presumably small) number of Boltzmann fluctuations in a landscape of vacua that includes a Minkowski vacuum.

This intuition can be bolstered by an analogy to one-dimensional quantum mechanics in the presence of a barrier. Consider once again a particle of mass m and

energy E moving in a potential $V(x)$ schematically similar to the false-vacuum potential shown in Figure 7.3. The particle can escape the well by tunneling through the barrier. A wave packet initially in the potential well will leak out, and the WKB approximation relates the wave functions on either side of the potential:

$$\frac{\psi(x_e)}{\psi(x_0)} = \exp\left(-\frac{1}{\hbar}\int_{x_0(E)}^{x_e(E)}\sqrt{2m(V(x)-E)}dx\right) \equiv e^{-\gamma/2} , \qquad (7.4.4)$$

where $x_0(E)$ and $x_e(E)$ are the starting and ending points for the region where the particle "has negative energy," so $V(x_0(E)) = V(x_e(E)) = E$. The escape probability is simply $e^{-\gamma}$, and the tunneling rate is given by the product of this probability with some characteristic frequency:

$$R = f(E)e^{-\gamma} . \qquad (7.4.5)$$

The classic barrier penetration problem considers a square-well potential, in which the bound particle has a position-independent momentum, $p(E) = \sqrt{2m(E-V)}$, and a characteristic "collision frequency", $f(E) = p(E)/(2mx_0)$. Here, we assume a more general potential, so the momentum is a function of both E and x, and the frequency will be given by some integral over positions inside the well. The exact expression is not important for us—we assume only that the frequency is an increasing function of E, $f'(E) > 0$. Then, the energy dependence of the tunneling rate is

$$\frac{dR}{dE} = f'(E)e^{-\gamma} - \frac{2}{\hbar}f(E)e^{-\gamma}\int_{x_0(E)}^{x_e(E)}\left[-\frac{2}{\sqrt{2m(V(x)-E)}}\right] , \qquad (7.4.6)$$

which is manifestly positive. (We have used the fact that $V(x_0)-E = V(x_e)-E = 0$ to eliminate the terms which arise from varying the limits of integration.)

This simple exercise demonstrates an intuitively sensible result: among states trapped behind a barrier, those with higher energy tunnel out more quickly. In the case of the cosmological false vacuum, the analogous statement is that excited states of the perturbative Hamiltonian undergo false-vacuum decay more rapidly.

In complementarity, we see that only in the case of a Minkowski true vacuum can recurrences and Boltzmann fluctuations be avoided entirely. A version of this phenomenon—the crucial difference in the long-term quantum evolution of landscapes with and without $\Lambda = 0$ vacua—has been previously noted in a slightly

different context [198, 208]. There, it was pointed out that quantum measurements in a false-vacuum state will decohere by becoming entangled with environment degrees of freedom, but they must eventually recohere if the total Hilbert space is finite-dimensional. In infinite-dimensional Hilbert spaces, in contrast, decoherence can persist forever. This argument is analogous to our own, in that such models are largely free of Boltzmann fluctuations.

7.5 Consequences

7.5.1 Boltzmann brains

In the conventional picture, because de Sitter space has a temperature, it experiences thermal fluctuations that lower the entropy by ΔS with a finite rate proportional to $e^{-\Delta S}$. If the Hartle-Hawking vacuum is eternal, then all dynamical fluctuations that fit within a horizon volume are produced an infinite number of times inside each such volume. Such fluctuations could contain conscious observers like ourselves [164–166, 179, 209–213]. Due to the exponential suppression of lower-entropy states, the fluctuations containing observers—even the ones that contain exact copies of our own brains—that occurred most frequently would look entirely unlike the world we observe. In particular, fluctuations containing the room you are reading this paper in would be vastly more likely than fluctuations containing all of Earth, let alone the entire observable universe, and the momentary coalescence of your brain thinking the precise thoughts you are having right now out of thermal equilibrium would be likelier still. If this conclusion were correct, we would not be able to trust our memories or our (supposed) observations, a solution inconducive to the practice of science.

We have argued, however, that this situation is less generic in de Sitter cosmologies than is often supposed.[10] The appearance of Boltzmann brains is avoided in the context of QFT in eternal de Sitter space or in a landscape with a terminal Minkowski vacuum (with or without complementarity). In these cases, the dimension of the Hilbert space is infinite, so the recurrence time also goes to infinity, and the (possibly false) de Sitter vacuum becomes quiescent. If the horizon volume is initially in an excited state (as it is if the dark energy is a positive cosmological constant), then the cosmic no-hair theorem dictates that correlations fall off exponentially with time as the excitations leave the horizon. The total number of Boltzmann brains will

---

[10]For related work that questions the validity of Boltzmann brains for decoherence-based reasons, see [180, 214, 215]. For the need for Hilbert space to be infinite-dimensional, see [216].

thus be finite and presumably small, given the vast exponential suppression of macroscopic fluctuations. Thus, if enough observers are produced before de Sitter space approaches the vacuum (*e.g.*, in a period of structure formation) the vast majority of observers can, in fact, trust their memories and observations. This conclusion opens the door for many multiverse models that might have been discounted because of a Boltzmann brain problem, and could help resolve potential tensions with low-energy physics [217].

7.5.2 Landscape eternal inflation

Another kind of fluctuation into a lower-entropy state that is often invoked in de Sitter cosmology is uptunneling from one de Sitter vacuum state to another one of higher energy [162, 163]. Processes such as this can be crucial for populating an entire landscape of vacua, starting from a state concentrated on any particular field value.

Uptunneling is conceptually very similar to the standard picture of a fluctuation into a Boltzmann brain: a vacuum in a thermal state undergoes a transition to a lower-entropy configuration with probability $e^{-\Delta S}$. The situation is the time-reverse of the well-known process of vacuum decay, which results in the production of particles and an increase in entropy. The analysis presented in this paper leads to an analogous conclusion to that of the last subsection: if the total Hilbert space is infinite-dimensional, excitations around any particular false vacuum will dissipate. As discussed in Section 7.4, the system will relax to a (perturbative, semi-perturbative, or true) vacuum state, not a state of definite field value. The state becomes quiescent, and the rate of Boltzmann fluctuations asymptotes to zero.

Note that eternal inflation is still conceivable: uptunneling is suppressed, but downtunneling proceeds as usual, and different branches of the wave function will correspond to different distributions of bubbles in a semiclassical spacetime background. If the field starts out in a metastable vacuum, then the portion of it that remains there (on any one branch) is rewarded with greater volume production. Almost every world line will intersect a bubble of lower-energy vacuum, but if the tunneling rate is low enough to avoid percolation, the physical volume remaining in the high-energy vacuum grows without bound, as depicted in the rightmost diagram in Figure 7.4. In this sense inflation continues forever.

On the other hand, it is clear that the details of eternal inflation in a landscape of vacua will change. In particular, the conclusions of the previous section suggest a reinterpretation of the rate equations for eternal inflation that relate the probabilities

of transitions between different vacua [218, 219, 279]. Consider the simple landscape of Figure 7.3, with minima located at field values φ_F and φ_T, respectively. In the standard presentation, *e.g.* [218], the rate equations for a two-minimum landscape read

$$\frac{dp_f}{d\tau} = -\kappa_f p_f + \kappa_t p_t \; , \qquad \frac{dp_t}{d\tau} = -\kappa_t p_t + \kappa_f p_f \; , \qquad (7.5.1)$$

where κ_f and κ_t are transition probabilities per unit proper time. The usual interpretation is that κ_f (κ_t) represents the probability to transition from the false (true) vacuum to the true (false) one. But we have argued that, in the long-time limit, the probability to transition from the true to a false vacuum falls to zero. However, both the true and the false vacuum states have nonzero overlap with the states of any definite field value, so heuristically we may think of the true vacuum, for example, as containing an exponentially small piece with field value near φ_F. The rate equations should essentially be interpreted as probabilities to transition between states of definite field value in an (unrealistic) idealization where an observer is measuring the value of the field at regular intervals. In the real universe, where there is no external observer and the wave function evolves unitarily, the state simply evolves toward the true vacuum as time passes. Dynamical fluctuations in de Sitter space do not provide a mechanism for populating an entire landscape with actual semiclassical geometries centered on different vacua and living on different branches of the wave function.

With horizon complementarity, this picture changes somewhat. If the true vacuum is de Sitter, Hilbert space is finite-dimensional, and Boltzmann fluctuations will lead to true transitions between states concentrated at different minima of the potential. If the true vacuum is Minkowski, on the other hand, Hilbert space is infinite-dimensional, and the above discussion is once again valid.

7.5.3 Inflationary perturbations

The absence of dynamical fluctuations in the de Sitter vacuum might seem to call into question the standard picture of the origin of density perturbations in inflation. In this case, however, the conventional wisdom gets the right answer; our approach leaves the standard predictions for density and tensor fluctuations from inflation essentially unaltered. The basic point is that the quantum state of light fields can remain coherent during inflation itself, and possess (non-dynamical) vacuum fluctuations, but then experience decoherence and branching of the wave function when entropy is generated at reheating.

We can describe the Hilbert space during inflation as a product of the quantum

states of the large-scale homogeneous background $\varphi(t)$ (macroscopic) perturbations and the small-scale (microscopic) perturbations:

$$\mathcal{H} = \mathcal{H}_{\varphi(t)} \otimes \mathcal{H}_{\text{macro}} \otimes \mathcal{H}_{\text{micro}} . \tag{7.5.2}$$

The small-scale perturbations, including the specific microstates of individual photons and other particles, are unobservable, in the same way that individual atoms and molecules are unobservable in an ordinary box of gas. They serve as an environment we can trace over to understand the state of the observable large-scale perturbations. During inflation, the overall quantum state approaches a factorizable form, as excitations dissipate and perturbations approach their lowest-energy states:

$$|\Psi_{\text{inflation}}\rangle = |\varphi(t)\rangle \otimes |0\rangle_{\text{macro}} \otimes |0\rangle_{\text{micro}} . \tag{7.5.3}$$

The state $|0\rangle_{\text{macro}}$ has a nonzero variance for the field operator φ, as calculated in standard treatments, but its quantum coherence is maintained.[11]

At reheating, however, entropy is generated. Energy in the inflaton is converted into a dense, hot plasma with many degrees of freedom. The specific form of the microscopic perturbations will depend on the state of the macroscopic perturbations; these factors become entangled, producing a state of the form

$$|\Psi_{\text{reheating}}\rangle = |\varphi(t)\rangle \otimes \left[\sum_i |\delta\varphi_i\rangle_{\text{macro}} \otimes \left(\sum_\mu |\delta\varphi_{i,\mu}\rangle_{\text{micro}} \right) \right] . \tag{7.5.4}$$

Tracing over the microscopic fluctuations leaves a mixed-state density matrix for the macroscopic fluctuations, inducing decoherence [220–225]. By this process, the unique quantum state of the inflaton field evolves into a large number of decohered branches, each with a specific pattern of perturbations such as we observe in the CMB, with statistics given by the Born rule. In effect, reheating acts as an explicit measurement process. We, therefore, expect that the standard calculations of scalar and tensor fluctuations in any given inflationary model are unaffected by the considerations in this paper.

---

[11]One might imagine that decoherence occurs because modes become super-Hubble-sized, and we should trace over degrees of freedom outside the horizon. This reasoning is not quite right, as such modes could (and often do) later re-enter the observable universe; they become larger than the Hubble radius during inflation but never leave the true horizon.

7.5.4 Stochastic eternal inflation

We next turn to the possibility of eternal inflation in a slow-roll potential, as distinguished from a landscape of false vacua. The traditional approach to this scenario makes use of the stochastic approximation, which treats the inflaton field value in the slow-roll regime as a stochastic variable, undergoing a random walk [154–156]; for recent treatments see [226–228]. Consider the case of a power-law potential,

$$V(\varphi) = \frac{\lambda \varphi^{2n}}{2n M_{\mathrm{pl}}^{2n-4}} \ . \tag{7.5.5}$$

In a single Hubble time, the expectation value of the field decreases by

$$\Delta\varphi = \frac{n M_{\mathrm{pl}}^2}{4\pi\varphi} \ , \tag{7.5.6}$$

but the dispersion of perturbations around this value is

$$\Delta^2 = \left\langle \delta\varphi^2 \right\rangle = \frac{H^3}{4\pi^2} t \ . \tag{7.5.7}$$

In a Hubble time H^{-1}, we have $\Delta = H/2\pi$.

   Now comes the critical step. In the stochastic approximation, one asserts that Δ represents an RMS fluctuation amplitude

$$|\delta\varphi| = \frac{H}{2\pi} \ , \tag{7.5.8}$$

and that the effective value of the inflaton in a given Hubble patch should be treated as a random variable drawn from a distribution with this amplitude. Above a critical field value,

$$\varphi^* = \lambda^{-1/(2n+2)} M_{\mathrm{pl}} \ , \tag{7.5.9}$$

the fluctuations dominate, $|\delta\varphi| \gg \Delta\varphi$. In this picture, to an excellent approximation, φ undergoes a random walk with time step H^{-1} and step size $|\delta\varphi|$. Causality dictates that each horizon area undergoes these fluctuations independently. Every Hubble time, when a horizon volume grows by a factor $e^3 \sim 20$, the field value in approximately 10 of the new horizon volumes is larger than its parent. In fact, this statement is a much stronger condition than required for eternal inflation. It suffices for only one of these volumes to move upward on the potential: $|\delta\varphi| \approx \mathcal{O}\left(\Delta\varphi/20\right)$.
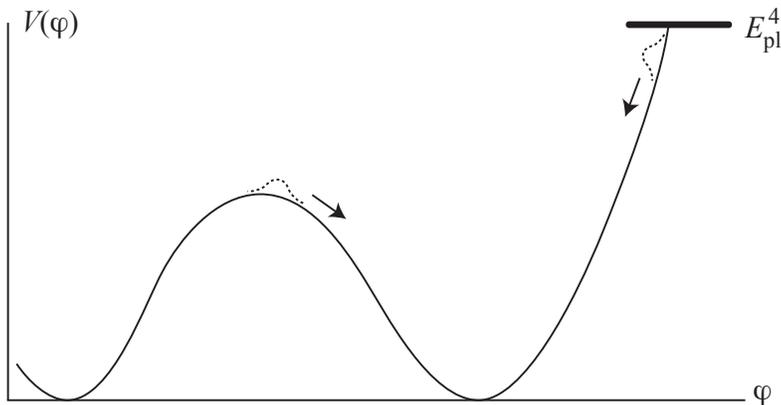
Figure 7.5. Potential supporting different kinds of inflation. Dashed lines are schematic representations of two different initial quantum states for the field. If the field begins at the right edge near the Planck cutoff, we expect it to evolve smoothly to the non-inflationary regime at the bottom of the potential. In contrast, if it begins at the top of a hill, it is plausible to imagine that part of the wave function remains in an inflationary state for arbitrarily long periods of time (although the amplitude for that branch of the wave function will be monotonically decreasing).

The stochastic-approximation approach to eternal inflation is in tension with the analysis presented in this paper. As we have argued in Section 7.2, quantum fluctuations in closed systems near equilibrium cannot be treated as classical random variables. Fluctuations $\delta\varphi$ only become real when they evolve into different decoherent branches of the wave function and generate entropy (what we have called measurement-induced fluctuations). For the perturbations we observe in the CMB, this entropy source is provided by reheating. But precisely in the slow-roll regime, where the stochastic inflation story is invoked, there is no entropy production, no measurement or decoherence, and no branching of the wave function. All that happens during a Hubble time is a decrease in the classical field expectation value, $\Delta\varphi$. There is no quantum-dominated regime; the field simply rolls down its potential.

A more honest approach to eternal inflation would be to take the quantum nature of the dynamics seriously, and investigate the evolution of the wave function describing the coupled background and perturbations; we hope to study this more carefully in future work. Nevertheless, it is possible to draw some qualitative conclusions by considering the evolution of a wave packet in field space representing the homogeneous mode. If the initial state of the field has support near a local maximum

of the potential, inflation is plausibly eternal: part of the wave packet will roll down the potential, eventually couple to perturbation modes, and experience decoherence, while part will remain near the maximum and continue to inflate. In contrast, if the field is slowly rolling down a monotonic portion of the potential—as expected for a polynomial potential with a Planck-scale energy density cutoff—it will reach the bottom of the potential, and the inflationary phase will end in a finite time and after a finite number of *e*-folds. These two possibilities are portrayed in Figure 7.5. We note that the simplest inflaton potentials, monomial power-laws $V(\varphi) \sim \varphi^n$, do not have saddle points and should thus avoid eternal inflation given a Planck-scale cutoff. (The recent BICEP2 detection [229] of large-scale B-mode polarization in the CMB, if interpreted as a tensor/scalar ratio $r \sim 0.2$, is well fit by an $m^2\varphi^2$ potential.)

For a field on the monotonic portion of the potential, one might object that, even once the field has rolled down, some portion of the wave function will always remain arbitrarily close to the maximum allowed value of the potential, *e.g.* the Planck-scale cutoff, just as a wave packet is supported throughout all of space despite being concentrated around a single point. This reasoning is correct, but it does not imply that there are some portions of the wave function where the end of inflation is postponed. The problem with this interpretation was already noted in Subsections 7.4.1 and 7.5.2: states of definite field value are not the same as states of definite energy density. In the slow-roll approximation, the cosmic no-hair theorem acts to bring the inflaton field to the appropriate vacuum state—a state of energy density corresponding perturbatively to de Sitter space with the appropriate cosmological constant. Each such state has nonzero overlap with the states of definite field value, but the cosmic no-hair theorem guarantees (to the extent that the slow-roll approximation is valid, so that no entropy is produced) that the field is driven into the appropriate false vacuum state, and then rolls smoothly to states with lower and lower energy density until the point that inflation ends. Again, there is negligible entropy production, no measurement-induced fluctuations, and no branching during this period—the inflaton remains in a single coherent state until reheating occurs.

To gain intuition for the points we make above, it is useful to consider applying the stochastic approximation to a free massive scalar field in eternal de Sitter space itself. At the minimum of the potential $V = m^2\varphi^2/2$, it is clear that the classical change $\Delta\varphi$ vanishes while the quantum variance $\delta\varphi$ does not, so the system is automatically in the quantum-dominated regime. If the stochastic approximation is applied, we expect occasional fluctuations of the field to very large values, leading to rapid inflation in those regions but not in others. In other words, if the stochastic

picture is valid, one is led to the conclusion that de Sitter space with a massive scalar field has a runaway instability, in contrast with the usual view that there is a lowest-energy eigenstate with a stable semiclassical geometry (*c.f.* [230]). In light of the above, we interpret this purported instability differently: it indicates a problem with the stochastic approximation, not with de Sitter space itself. The vacuum state of the scalar field is not a state of definite field value, although it is centered around the minimum of the potential. Rather, the state has overlap with all field values, at least up to a potential Planck-scale energy density cutoff. But we do not interpret the de Sitter vacuum as an unstable superposition of different field values expanding at different rates. Instead, we say that the field is in a single state, the vacuum, with a definite energy density given by the cosmological constant Λ.

7.5.5 Other formulations of quantum mechanics

Throughout this paper we have worked in the context of the Everett/Many-Worlds formulation of quantum mechanics, in which a single wave function evolves unitarily in Hilbert space according to the Schrödinger equation. Our conclusions could be dramatically altered in other formulations. We will not explore these possibilities in detail here, but we briefly mention two alternatives.

One would be the de Broglie-Bohm approach, in which the variables include both a wave function and variables in a separate configuration space [231–234]. In such a theory, the wave function could be completely stationary (as in the de Sitter vacuum), but the configuration-space variables could still fluctuate. What we think of as a stationary thermal state in the Everett approach would be more closely analogous to a thermal distribution function in classical statistical mechanics; while the density operator is stationary, the underlying state could still be evolving in time. We might, therefore, observe dynamical fluctuations even in equilibrium. Recent work has argued, however, that in practice these Boltzmann fluctuations can be avoided in Bohmian cosmology [235].

Another alternative is a stochastic dynamical-collapse theory, such as the Ghirardi-Rimini-Weber (GRW) model [236, 237]. Set in the context of nonrelativistic, many-particle quantum mechanics, the wave function has a fixed probability per particle per unit time of spontaneously collapsing to a localized position. Entanglement between particles induces an effective, ongoing "measurement" of macroscopic systems. There is not a well-developed GRW model for QFT in de Sitter space, but the philosophy of the approach leads us to expect that a thermal state would experience true fluctuations; the possibility of dynamical collapse introduces a new kind of time-

dependence that would be absent in equilibrium in Everettian quantum theory. It seems we are dealing with one of the rare cases in which one's favorite formulation of quantum mechanics can drastically affect one's expectation for how observable quantities evolve.

7.6 Conclusions

Quantum variables are not equivalent to classical stochastic variables. They can be related by the appearance of measurement-induced fluctuations, which require entropy generation, decoherence, and branching of the wave function. In stationary states, entropy is not generated, and the wave function remains fixed; therefore, there are no dynamical fluctuations, and treating a quantum field as a classical stochastic field is inappropriate. We have argued that this shift in thinking has important consequences for the cosmology of de Sitter space, since de Sitter regions tend to approach a stationary thermal state. In particular, if the true Hilbert space is infinite-dimensional (as is the case in QFT in curved spacetime or in horizon complementarity in the presence of a Minkowski vacuum), de Sitter vacua settle down and do not fluctuate. There are no Boltzmann brains in such states, relieving a major problem for many multiverse cosmological models. On the other hand, we also suggest there is neither uptunneling to higher-energy vacua nor stochastic fluctuations up a slow-roll potential, implying that eternal inflation is much less generic than often supposed. A better understanding of complementarity and the correct formulation of quantum mechanics will help establish what happens in the real universe.

Acknowledgments

How Decoherence Affects the Probability of Slow-Roll Eternal Inflation

8.1 Introduction

The state of the early universe – hot, dense, and very smooth – is extremely fine-tuned by conventional dynamical measures [238]. Inflationary cosmology [239–241] attempts to account for this apparent fine-tuning by invoking a period of accelerated expansion in the very early universe. The potential energy of a slowly rolling scalar field, the inflaton, serves as a source of quasi-exponential expansion through the Friedmann equation, leading to a universe that is nearly smooth and spatially flat.

Quantum mechanics, however, changes this picture of slow-roll inflation in an important way. Although the classical equations of motion completely determine the behavior of the inflaton zero mode (i.e. the expectation value of the field) rolling down the potential, quantum field theory in curved spacetime dictates that each Fourier mode of the field has a nonzero variance (two-point function). This variance persists after a mode leaves the Hubble radius and classically freezes out, and it is still present when inflation ends and the mode re-enters the Hubble radius. If reheating at the end of inflation produces a sufficiently rich thermal bath of particles and radiation, decoherence [3–7] occurs (if it has not already): the thermal bath becomes entangled with definite values of the curvature perturbation entering the Hubble radius, so that the quantum states corresponding to different values of the inflaton field become orthogonal and evolve without interference [220–225, 242, 243]. Hence, any modes within a Hubble volume after the end of inflation have inevitably undergone decoherence; our observable universe, including the Cosmic Microwave Background (CMB) and large-scale structure, is one branch of the universal quantum state.

Slow-roll *eternal* inflation occurs when there is a period during which the quantum variance in the inflaton field is sufficiently large that the field may fluctuate upward on its potential [226, 228, 244–246]. In regions where these upward fluctuations occur, the universe expands at a faster rate, and such regions come to dominate the physical volume of space. If the probability of upward fluctuations is sufficiently high, the total volume of inflating space expands as a function of time, and inflation is eternal. Although there are other mechanisms to achieve an eternally inflating universe, such as tunneling transitions which produce inflating bubbles [154], we concentrate on slow-roll eternal inflation and refer to it simply as eternal inflation throughout the paper.

Eternal inflation hinges on the idea that quantum fluctuations of the inflaton are true, dynamical occurrences. However, quantum fluctuations become dynamical in

unitary (Everettian, Many-Worlds) quantum mechanics only when decoherence and branching of the wave function occur [247]. To put the slow-roll eternal inflation story on a firm foundation, it is therefore necessary to examine carefully just when inflationary modes decohere, and how that decoherence enables backreaction that can effect the value of the expansion rate in different regions.

In this paper we therefore investigate eternal inflation carefully from a quantum-mechanical perspective. Following the approach of the recent work of Ref. [248], we work with the adiabatic curvature perturbation ζ and consider the lowest-order gravitationally-sourced interaction between modes of different wavelengths. This interaction vanishes in the limit as slow-roll parameters go to zero, and therefore maintains the stability of pure de Sitter space itself, where no decoherence should occur [247]. It was shown in Ref. [248] that this interaction decoheres the modes that we observe in the CMB on $\mathcal{O}(10)$ Hubble times after they cross the Hubble radius. We consider the effects of this long-wavelength decoherence on the evolution of modes that still have short wavelengths compared to the Hubble radius at the time of decoherence, which we use as a proxy for the cosmological backreaction due to the decoherence. We find that the standard lore in which eternal inflation occurs when quantum dispersion dominates over classical rolling down the potential is qualitatively correct, but we also show that the quantitative predictions of eternal inflation must be adjusted to incorporate the time it takes for gravitational interactions to bring about decoherence.

The remainder of this paper is structured as follows. In Section 8.2 we review the standard picture of slow-roll eternal inflation and explain the basic quantum-mechanical picture behind our analysis. In the next two sections we construct the technical machinery needed to establish the details of our picture of eternal inflation. In Section 8.3 we set up the general problem of finding the time evolution of the inflaton field and describe its solution by path-integral methods and Feynman diagrams. We review the result of Ref. [248] that gravitational backreaction decoheres super-Hubble adiabatic curvature modes during inflation. In Section 8.4 we interpret this result in the language of wave function branching, and introduce the notion of observables within a particular branch, where the long-wavelength decohered modes have a definite classical value. We describe the Feynman rules for computing these observables, and show in particular that the evolution of short-wavelength modes depends on the long-wavelength background, suggesting that different decohered branches have different cosmological histories. In Section 8.5 we then use this machinery to study eternal inflation. We consider the statistics of the daughter cos-

mologies that emerge from a single region of space as super-Hubble modes decohere and the wave function branches. We write the probability of the effective upward evolution of the cosmological constant that heralds eternal inflation as a function of the inflationary potential. The expression for the probability, as expected, largely reproduces previous results, with slight modifications as a result of correctly incorporating a potential-dependent time until decoherence. Finally, we discuss the broader implications of this work for the standard eternal inflation in Section 8.6 and then conclude in Section 8.7.

8.2   The Basic Picture

To set the stage, let us consider this picture more closely. In order to determine the global structure of a universe in which inflation has begun, it is necessary to consider modes which have left the Hubble radius and have yet to return—and indeed will possibly never return, due to the present acceleration of the universe. If super-Hubble modes decohere in some particular basis, the quantum state of the universe as a whole can be written as a superposition of different states with definite values of the modes in that basis—"branches"—which do not interfere with one another. In particular, some branches may have definite values of cosmological parameters, such as the Hubble constant, which differ from the values on the initial classical slow-roll trajectory. Although the expectation values themselves will not change, individual classical patches after inflation may have values of the parameters that differ strongly from the expectation values. Even if the parameters of a particular inflationary potential are chosen to produce a particular amplitude $\delta\rho/\rho$ for the density perturbations, for example, some of the classical cosmologies resulting from inflation on this potential will nevertheless have entirely different values. If decoherence produces a distribution of Hubble constants around the classical value, there will be some branches of the wave function on which the Hubble constant grows rather than decreases monotonically according to the equations of motion and hence on which the end of inflation can be postponed indefinitely. If these branches are common enough, the volume of inflating space may grow indefinitely. There is no global spacelike hypersurface on which inflation ends, and the universe is in the regime of eternal inflation [226].

It is therefore important to understand if eternal inflation actually occurs and under what conditions. In the standard picture of inflation, the Hubble rate of

expansion is determined by

$$\frac{\dot{a}}{a} = \sqrt{\frac{V(\phi)}{3}} = H(\phi) \; , \tag{8.2.1}$$

where $8\pi G = c = \hbar = 1$, the dot notation indicates a derivative with respect to the physical time t, and $\phi \equiv \langle\phi\rangle + \delta\phi$ is the inflaton field. Quantum fluctuations of $\delta\phi$ behave as [244, 245, 249]

$$\langle\delta\phi^2(t+\Delta t)\rangle - \langle\delta\phi^2(t)\rangle = \frac{H^3}{4\pi^2}\Delta t \tag{8.2.2}$$

over a time Δt. According to the standard story, the quantum state of a mode collapses when it reaches the Hubble scale – corresponding in our language to decoherence – and each mode obtains a value given by the sum of its classical evolution plus a quantum fluctuation up its potential [244, 245]. In the stochastic approximation, these super-Hubble modes are assumed to decohere quickly, and the evolution of the inflaton field is treated as a random walk on top of its classical slow-roll trajectory [154, 226, 245, 249]. In a Hubble time $\Delta t \sim H^{-1}$, the fluctuation in field value is $\Delta\phi \sim H/(2\pi)$. If the size of these fluctuations are sufficiently large, inflation may persist due to the scalar field stochastically fluctuating up in its potential, countering the classical motion. We will discuss this more extensively in Section 8.5 below.

The assumption of rapid decoherence does not necessarily hold in all circumstances, in which case eternal inflation must be treated appropriately in the context of quantum mechanics. Let us therefore be a bit more explicit about the relationship between backreaction and decoherence, in a simplified toy-model context.

Consider a Hilbert space decomposed into two factors $\mathcal{H} = \mathcal{H}_L \otimes \mathcal{H}_S$, corresponding roughly to long-wavelength and short-wavelength modes. Let $\{|\phi_i\rangle\}$ be a basis for $\mathcal{H}_L$ and $\{|\omega_a\rangle\}$ be a basis for $\mathcal{H}_S$. We would like to illustrate the relationship between entanglement and backreaction. Therefore consider a state of the form

$$|\Psi\rangle = \alpha|\phi_1\rangle|\omega_1\rangle + \beta|\phi_2\rangle|\omega_2\rangle \; . \tag{8.2.3}$$

For generic α and β such a state is clearly entangled, but for $\alpha = 1$, $\beta = 0$ it is a product state, so this form suffices to examine both possibilities.

We would like to illustrate the (perhaps intuitive) fact that the evolution of the short-wavelength states can depend on that of the long-wavelength states with which they are entangled, but without entanglement it will simply depend on the long-

wavelength state as a whole. In the absence of entanglement (and the decoherence that leads to it) there are no fluctuations or quantum jumps; in particular, it does not matter if the form of that state is that of a squeezed state [250, 251].

We therefore consider an interaction Hamiltonian that does not itself lead to decoherence; in other words, one that is a tensor product of operators on the two factors of Hilbert space, $\hat{H}^I = \hat{h}^{(L)} \otimes \hat{k}^{(S)}$. The matrix elements of such a Hamiltonian in the $\{|\phi_i\rangle, |\omega_a\rangle\}$ basis take the form

$$H^I_{iajb} = h^{(L)}_{ij} k^{(S)}_{ab} \ . \tag{8.2.4}$$

Its action on the state (8.2.3) is

$$\hat{H}^I|\Psi\rangle = \alpha \sum_{jb} \left(h^{(L)}_{1j}|\phi_j\rangle \right) \otimes \left(k^{(S)}_{1b}|\omega_b\rangle \right) + \beta \sum_{jb} \left(h^{(L)}_{2j}|\phi_j\rangle \right) \otimes \left(k^{(S)}_{2b}|\omega_b\rangle \right) \ . \tag{8.2.5}$$

From this form it should be clear that the evolution of the short-wavelength modes depends on the branch of the wave function they are in. In the α branch they evolve under the influence of the components $k^{(S)}_{1b}$, while in the β branch they evolve under the influence of $k^{(S)}_{2b}$. If the state were unentangled, there would be no differentiation in how different parts of the long-wavelength state might affect the evolution of the shorter modes. In this way, decoherence is necessary for backreaction to occur differently within different branches. It is therefore important to examine the rate of decoherence during inflation to accurately calculate the stochastic evolution of the inflationary spacetime on each branch.

8.3 Gravitational Decoherence during Inflation

We would like to understand the full quantum dynamics of the inflaton field during slow-roll inflation. Following [248], we write down an expression for the wave function and then extract information about particular modes of interest. We confine ourselves in this section and the next to perturbative quantum field theory in curved spacetimes rather than full nonperturbative quantum gravity, so we carry out the calculations on a fixed de Sitter background. We argue below that our perturbative results, when appropriately interpreted, nevertheless suffice to determine how backreaction alters the effective Hubble constant and hence determine when eternal inflation occurs. Since we are tracking the evolution of the wave function, we work in the Schrödinger picture rather than in the interaction picture used in typical flat-

space QFT calculations: we view states rather than operators as evolving in time, and our expectation values are always with respect to the wave function at the time of interest rather than S-matrix elements.

8.3.1 The general problem

We want to consider the (coordinate or conformal) time evolution of (particular modes of) a quantum state $|\Psi\rangle$ in the Hilbert space $\mathcal{H}_\zeta$ of a quantum field theory of a single real scalar field ζ with translationally and rotationally invariant interactions. A natural basis spanning this Hilbert space is the basis of field configurations, which we can think of either as functions of position space $\zeta(\mathbf{x})$ or, more often, as functions of momentum space $\zeta(\mathbf{k})$. We decompose $\mathcal{H}_\zeta$ into an infinite tensor product of factors representing each point in (position or momentum) space,

$$\mathcal{H}_\zeta = \bigotimes_{\mathbf{k}} \mathcal{H}_{\zeta,\mathbf{k}} \ , \tag{8.3.1}$$

so that a particular field configuration $|\zeta\rangle$ is the product of a specific multi-particle state in each individual Hilbert space factor,

$$|\zeta\rangle = \bigotimes_{\mathbf{k}} |\zeta_{\mathbf{k}}\rangle \ . \tag{8.3.2}$$

Each $|\zeta_{\mathbf{k}}\rangle$ is an eigenstate of the field value operator $\hat{\zeta}_{\mathbf{k}}$ on the appropriate factor $\mathcal{H}_{\zeta,\mathbf{k}}$:

$$\hat{\zeta}_{\mathbf{k}}|\zeta_{\mathbf{k}}\rangle = \zeta_{\mathbf{k}}|\zeta_{\mathbf{k}}\rangle \ . \tag{8.3.3}$$

Thus a field configuration $|\zeta\rangle$ is a simultaneous eigenstate of all operators which consist of the tensor product of the field value operator in a given Hilbert space factor $\mathcal{H}_{\zeta,\mathbf{k}}$ and the identity in all other factors. The collection of all of the eigenvalues $\zeta_{\mathbf{k}}$ comprises the field configuration as a function of momentum space, $\zeta(\mathbf{k})$.

Given this basis, it is often convenient to work with the wave functional $\Psi[\zeta]$ instead of the state itself:

$$\Psi[\zeta] \equiv \langle\zeta|\,\Psi\rangle \ . \tag{8.3.4}$$

We work in the Schrödinger picture and consider states rather than operators as evolving in time. Time evolution is generated by the Hamiltonian $\hat{H}(t)$; the symmetry assumptions mean that can we decompose it as a sum of symmetry-respecting polynomial interactions among the fields $\zeta_{\mathbf{k}}$ and the canonical momenta $\pi_{\mathbf{k}}^{(\zeta)} \equiv -i(\delta/\delta\zeta_{-\mathbf{k}})$. The lowest-order terms, up to quadratic order in the fields, make up

the free Hamiltonian $\hat{H}_{\text{free}}$. Given $\hat{H}_{\text{free}}$, we can write a special Gaussian state $|\Psi_G\rangle$, which is the superposition of field configuration basis states with coefficients given by the weight $\Psi_G[\zeta](t)$ that solves the Schrödinger equation:

$$|\Psi_G\rangle = \sum_\zeta \Psi_G[\zeta]|\zeta\rangle, \qquad i\frac{d}{dt}\Psi_G[\zeta] = \hat{H}_{\text{free}}[\zeta](t)\Psi_G[\zeta] . \tag{8.3.5}$$

This weight is given by a Gaussian integral over the field modes:

$$\Psi_G[\zeta](t) \equiv N_\zeta(t) \exp\left[-\int_{\mathbf{k}} \zeta_{\mathbf{k}} \zeta_{\mathbf{k}}^\dagger A_\zeta(k,t)\right] , \tag{8.3.6}$$

where $N_\zeta(t)$ is a normalization constant, the shorthand notation for the integral is given by Eq. (8.A.4) below, the complex conjugate (denoted with †) enforces the reality condition on $\zeta(\mathbf{x})$, and A_ζ depends only on the magnitude of k by the symmetry assumption. The function $A_\zeta(k,t)$ is given implicitly by Eq. (8.3.5), and we derive it explicitly for our Hamiltonian of interest below.

We assume that the initial (at $t = 0$ or equivalently $\tau = -\infty$) state is simply

$$\Psi[\zeta](t = 0) = \Psi_G[\zeta](t = 0) . \tag{8.3.7}$$

Our assumption is motivated by the fact that this state has the form of the Euclidean[1] vacuum [185, 188–192], the unique state which is both de Sitter-invariant and well-behaved at short distances, i.e. obeys the Hadamard condition [187]. Nevertheless, it is an assumption: it implies in particular that short-wavelength modes which have just crossed the Planck scale and entered the domain of validity for QFT are in their vacuum state and unentangled with modes of different wavelengths.

8.3.2 The free action

We now specialize to the case of interest: perturbations around a de Sitter background. The background de Sitter metric in a flat slicing is $ds^2 = -dt^2 + a(t)^2\,d\mathbf{x}^2$, where $a(t) = e^{Ht} = -1/H\tau$. Concentrating solely on scalar modes, we work in a gauge in which fluctuations are represented as perturbations ζ of the induced spatial metric,

$$g_{ij} = a(t)^2 e^{2\zeta(\mathbf{x},t)} . \tag{8.3.8}$$

---

[1]The Euclidean vacuum is also known as the Bunch-Davies vacuum [145, 146] for a massive, noninteracting scalar field or the Hartle-Hawking vacuum [147] for an interacting one.

This curvature perturbation describes the amount of expansion at any point; if $\zeta \ll 1$, it describes the expansion in the given region. The quadratic action for ζ is

$$S_{\text{free}}[\zeta] = \frac{1}{2} \int d^4x \frac{2\epsilon M_p^2}{c_s^2} a^3 \left[\dot{\zeta}^2 - \frac{c_s^2}{a^2}(\partial_i\zeta)^2 \right] , \tag{8.3.9}$$

where $M_p \equiv 1/\sqrt{8\pi G}$ is the reduced Planck mass and $\epsilon \equiv -\dot{H}/H^2 \ll 1$ is the first slow-roll parameter. We set the propagation speed to $c_s = 1$; Appendix B of Ref. [248] treats the general case. We work in Fourier space, using the conventions in Appendix 8.A. Note that because $\zeta(\mathbf{x}, t)$ is real we have $\zeta_{\mathbf{k}}^\dagger = \zeta_{-\mathbf{k}}$, at least classically. It is also true quantum-mechanically if the quantum state is invariant under $\mathbf{k} \to -\mathbf{k}$, which is the case for our initial vacuum state. In Appendix 8.A we use the free action (8.3.9) to derive the free Hamiltonian

$$\hat{H}_{\text{free}}[\zeta] = \frac{1}{2} \int_{\mathbf{k}} \left[\frac{1}{2\epsilon M_p^2 a^3} \pi_{\mathbf{k}}^{(\zeta)} \pi_{-\mathbf{k}}^{(\zeta)} + 2\epsilon M_p^2 a k^2 \zeta_{\mathbf{k}} \zeta_{-\mathbf{k}} \right] , \tag{8.3.10}$$

and hence an expression for A_ζ,

$$A_\zeta(k, \tau) = k^3 \frac{\epsilon M_p^2}{H^2} \frac{1 - \frac{i}{k\tau}}{1 + k^2\tau^2} . \tag{8.3.11}$$

8.3.3   Interactions

Thus far we have worked only with the free Hamiltonian $\hat{H}_{\text{free}}[\zeta]$. The full Hamiltonian consists of the free term and an interaction term: $\hat{H}[\zeta] = \hat{H}_{\text{free}}[\zeta] + \hat{H}_{\text{int}}[\zeta]$. If the interaction Hamiltonian is perturbatively small, evolution with the full Hamiltonian instead of the free one adds an extra multiplicative term to the wave functional:

$$\Psi[\zeta](t) = \Psi_G[\zeta](t) \times \Psi_{NG}[\zeta](t). \tag{8.3.12}$$

The lowest-order interaction is cubic, so the non-Gaussian factor can be written

$$\Psi_{NG}[\zeta](t) \equiv \exp\left[\int_{\mathbf{k},\mathbf{k}',\mathbf{q}} \zeta_{\mathbf{k}} \zeta_{\mathbf{k}'} \zeta_{\mathbf{q}} \mathcal{F}_{\mathbf{k},\mathbf{k}',\mathbf{q}}(t) \right] , \tag{8.3.13}$$

where the shorthand notation for the integral, which includes a momentum-conserving delta function, is given by Eq. (8.A.4) below. Because we have taken $\hat{H}_{\text{int}}$ to be rotationally invariant, $\mathcal{F}_{\mathbf{k},\mathbf{k}',\mathbf{q}}$ depends only the magnitudes k, k', and q of the momenta.

We solve for $\mathcal{F}$ by writing the Schrödinger equation using $H[\zeta]$ and then sub-

tracting the free Schrödinger equation. Intuitively, $\mathcal{F}(\tau)$ represents the cumulative effect of all three-point interactions from the initial (conformal) time τ_0 to time τ. Each specific interaction is computed by using the free Hamiltonian to evolve up to an intermediate time τ', then inserting the interaction term at that time; the full effect is the result of integrating over all these intermediate times. The result is

$$\mathcal{F}_{\mathbf{k},\mathbf{k}',\mathbf{q}}(t) = i \int_{\tau_0}^{\tau} \frac{d\tau'}{H\tau'} \tilde{\mathcal{H}}_{\mathbf{k},\mathbf{k}',\mathbf{q}}^{(\text{int})}(\tau) \exp\left[-i \int_{\tau'}^{\tau''} d\tau'' \alpha_{\mathbf{k},\mathbf{k}',\mathbf{q}}(\tau'')\right] , \qquad (8.3.14)$$

where $\tilde{\mathcal{H}}^{(\text{int})}$ is a classical source, defined implicitly through the action of $\hat{H}_{\text{int}}$ on Ψ_G,

$$\hat{H}_{\text{int}}[\zeta](t)\Psi_G[\zeta](t) \equiv \left[\int_{\mathbf{k},\mathbf{k}',\mathbf{q}} \zeta_{\mathbf{k}}\zeta_{\mathbf{k}'}\zeta_{\mathbf{q}} \tilde{\mathcal{H}}_{\mathbf{k},\mathbf{k}',\mathbf{q}}^{(\text{int})}(t)\right] \Psi_G[\zeta](t) . \qquad (8.3.15)$$

The quantity α implements the free evolution,

$$\alpha_{\mathbf{k},\mathbf{k}',\mathbf{q}}(\tau) \equiv \left[f_\zeta(k,\tau)A_\zeta(k,\tau) + f_\zeta(k',\tau)A_\zeta(k',\tau) + f_\zeta(q,\tau)A_\zeta(q,\tau)\right]/(H\tau), \qquad (8.3.16)$$

where f_ζ is the coefficient of the kinetic term in $H_{\text{free}}[\zeta]$,

$$f_\zeta(k,\tau) = \frac{1}{2\epsilon M_p^2 a^3} = -\frac{\tau^3 H^3}{2\epsilon M_p^2} . \qquad (8.3.17)$$

Note that $\mathcal{F}_{\mathbf{k},\mathbf{k}',\mathbf{q}}$ is completely symmetric in its three momentum arguments.

The physically relevant interaction term for the case of interest here is the gravitationally sourced $\zeta\zeta\zeta$ interaction which contains no time derivatives and hence does not vanish in the super-Hubble limit, where $\dot{\zeta}$ terms are redshifted away. We have defined ζ as the fluctuations around a de Sitter background, so the interaction terms should vanish in the limit of pure de Sitter space, i.e. they should have coefficients proportional to the slow roll parameters ϵ and η. In particular, the interaction Hamiltonian is [248, 252]

$$\hat{H}_{\text{int}}[\zeta] = \frac{M_p^2}{2} \int d^3\mathbf{x}\epsilon(\epsilon+\eta) a\zeta^2\partial^2\zeta . \qquad (8.3.18)$$

This expression for $\hat{H}_{\text{int}}$ then sets the form of $\mathcal{F}$; the computation is performed in Ref. [248], which finds in particular that in the late-time limit $\tau \to 0$ the imaginary part of $\mathcal{F}$ dominates, $|\text{Re}\,\mathcal{F}| \ll |\text{Im}\,\mathcal{F}|$. This means Ψ_{NG} can be approximated as a pure phase, $|\Psi_{NG}[\zeta](t)|^2 \approx 1$.

8.3.4 Feynman rules

In order to address the issue of backreaction, it is necessary to extend the results of Ref. [248] by going beyond the pure-phase approximation. Given the expression in Eq. (8.3.12), we can proceed to calculate expectation values of observables. In particular, we are interested in the evolution of short-wavelength, sub-Hubble modes. The free evolution of a mode is given by Eq. (8.3.11), which appears in the computation of the two-point function $\langle \zeta_{\mathbf{k}_\star} \zeta_{-\mathbf{k}_\star} \rangle$.

We begin by converting the operator expectation value into a path integral. For convenience we write the path integral over field configurations as $\int \mathcal{D}\zeta \equiv \int_\zeta$. Inserting a complete set of states with a definite field value in each momentum mode, we have

$$
\begin{aligned}
\langle \zeta_{\mathbf{k}_\star} \zeta_{-\mathbf{k}_\star} \rangle &= \left\langle \Psi \middle| \hat{\zeta}_{\mathbf{k}_\star} \hat{\zeta}_{-\mathbf{k}_\star} \middle| \Psi \right\rangle \\
&= \left\langle \Psi \middle| \left(\int_\zeta |\zeta\rangle\langle\zeta| \right) \hat{\zeta}_{\mathbf{k}_\star} \hat{\zeta}_{-\mathbf{k}_\star} \left(\int_{\zeta'} |\zeta'\rangle\langle\zeta'| \right) \middle| \Psi \right\rangle \\
&= \int_\zeta \Psi^\dagger[\zeta] \zeta_{\mathbf{k}_\star} \zeta_{-\mathbf{k}_\star} \Psi[\zeta] \qquad\qquad (8.3.19) \\
&= \frac{1}{N} \int_\zeta \zeta_{\mathbf{k}_\star} \zeta_{-\mathbf{k}_\star} \Psi_G^\dagger \Psi_G \Psi_{NG}^\dagger \Psi_{NG} \;. \qquad (8.3.20)
\end{aligned}
$$

To lowest order in $\operatorname{Re}\mathcal{F}/\operatorname{Im}\mathcal{F}$, Ψ_{NG} is a pure phase, so $\Psi_{NG}^\dagger \Psi_{NG} \approx 1$ and the path integral becomes Gaussian:

$$
\begin{aligned}
\langle \zeta_{\mathbf{k}_\star} \zeta_{-\mathbf{k}_\star} \rangle &\approx \frac{1}{N} \int_\zeta \zeta_{\mathbf{k}_\star} \zeta_{-\mathbf{k}_\star} \Psi_G^\dagger \Psi_G \\
&= \frac{\int_\zeta \zeta_{\mathbf{k}_\star} \zeta_{-\mathbf{k}_\star} \exp\left\{ - \int_{\mathbf{k}} \zeta_{\mathbf{k}} \zeta_{\mathbf{k}}^\dagger \left[A_\zeta^\dagger(k,t) + A_\zeta(k,t) \right] \right\}}{\int_\zeta \exp\left\{ - \int_{\mathbf{k}} \zeta_{\mathbf{k}} \zeta_{\mathbf{k}}^\dagger \left[A_\zeta^\dagger(k,t) + A_\zeta(k,t) \right] \right\}} \qquad (8.3.21) \\
&= \frac{(2\pi)^3 \delta^3(0)}{4 \operatorname{Re} A_\zeta\left(\mathbf{k}_\star, t\right)} \;, \qquad\qquad (8.3.22)
\end{aligned}
$$

recovering the free evolution[2]. Recall again that we are working in the Schrödinger picture, where the time dependence lives in the state $|\Psi\rangle$ rather than the operators, so the details of the calculation differ from the more familiar computation of the 2-point correlator from the path integral in QFT (though it should give the same

---

[2]Our expression differs by a factor of 2 from that in Eqs. (4.8-9) of Ref. [248], but as noted in Appendix 8.A our definition of A_ζ itself also differs by a factor of 2 and the two factors cancel here.

result); in particular note that because of the $\Psi^\dagger\Psi$ term it is not the action S itself but rather $S + S^\dagger = 2\,\mathrm{Re}\,S$ that appears in the exponential.

We see that the pure phase assumption ensures that the (even-point) correlation functions are unchanged by the interactions. Thus, to capture the effect of these interactions, we need to go beyond the pure phase assumption by writing the full expression for $\Psi_{NG}^\dagger\Psi_{NG}$ rather than simply approximating it as 1. We find

$$
\begin{aligned}
\Psi_{NG}^\dagger\Psi_{NG} &= \exp\left[\int_{\mathbf{k},\mathbf{k}',\mathbf{q}} \zeta_{\mathbf{k}}\zeta_{\mathbf{k}'}\zeta_{\mathbf{q}}\mathcal{F}_{\mathbf{k},\mathbf{k}',\mathbf{q}} + \int_{\mathbf{k},\mathbf{k}',\mathbf{q}} (\zeta_{\mathbf{k}}\zeta_{\mathbf{k}'}\zeta_{\mathbf{q}}\mathcal{F}_{\mathbf{k},\mathbf{k}',\mathbf{q}})^\dagger\right] \\
&= \exp\left[\int_{\mathbf{k},\mathbf{k}',\mathbf{q}} \zeta_{\mathbf{k}}\zeta_{\mathbf{k}'}\zeta_{\mathbf{q}}\mathcal{F}_{\mathbf{k},\mathbf{k}',\mathbf{q}} + \int_{\mathbf{k},\mathbf{k}',\mathbf{q}} \zeta_{-\mathbf{k}}\zeta_{-\mathbf{k}'}\zeta_{-\mathbf{q}}\mathcal{F}_{\mathbf{k},\mathbf{k}',\mathbf{q}}^\dagger\right] \\
&= \exp\left[\int_{\mathbf{k},\mathbf{k}',\mathbf{q}} \zeta_{\mathbf{k}}\zeta_{\mathbf{k}'}\zeta_{\mathbf{q}}\left(\mathcal{F}_{\mathbf{k},\mathbf{k}',\mathbf{q}} + \mathcal{F}_{\mathbf{k},\mathbf{k}',\mathbf{q}}^\dagger\right)\right] &\text{(8.3.23)} \\
&= \exp\left[\int_{\mathbf{k},\mathbf{k}',\mathbf{q}} 2\zeta_{\mathbf{k}}\zeta_{\mathbf{k}'}\zeta_{\mathbf{q}}\mathrm{Re}\,\mathcal{F}_{\mathbf{k},\mathbf{k}',\mathbf{q}}\right] . &\text{(8.3.24)}
\end{aligned}
$$

To obtain Eq. (8.3.23), we substitute $\mathbf{k}, \mathbf{k}', \mathbf{q} \to -\mathbf{k}, -\mathbf{k}', -\mathbf{q}$ in the second integrand, which leaves the integral unchanged, keeping in mind that $\mathcal{F}_{\mathbf{k},\mathbf{k}',\mathbf{q}}$ depends only on the magnitude of the momenta. As desired, the imaginary part of $\mathcal{F}$ drops out entirely, and the integrand vanishes in the limit $\mathrm{Re}\,\mathcal{F} \to 0$.

We now insert our improved expression for $\Psi_{NG}^\dagger\Psi_{NG}$ into the two-point function $\langle\zeta_{\mathbf{k}_\star}\zeta_{-\mathbf{k}_\star}\rangle$ (8.3.20):

$$
\langle\zeta_{\mathbf{k}_\star}\zeta_{-\mathbf{k}_\star}\rangle = \frac{1}{N}\int_\zeta \zeta_{\mathbf{k}_\star}\zeta_{-\mathbf{k}_\star}\exp\left[-\int_{\mathbf{k}} 2\zeta_{\mathbf{k}}\zeta_{\mathbf{k}}^\dagger\,\mathrm{Re}\,A_\zeta(k,t)\right]\exp\left[\int_{\mathbf{k},\mathbf{k}',\mathbf{q}} 2\zeta_{\mathbf{k}}\zeta_{\mathbf{k}'}\zeta_{\mathbf{q}}\,\mathrm{Re}\,\mathcal{F}_{\mathbf{k},\mathbf{k}',\mathbf{q}}\right] .
$$
$$
\text{(8.3.25)}
$$

Since we cannot integrate this expression analytically, we Taylor-expand the interaction term, assuming that each term in the integral is perturbatively small:

$$
\exp\left[\int_{\mathbf{k},\mathbf{k}',\mathbf{q}} 2\zeta_{\mathbf{k}}\zeta_{\mathbf{k}'}\zeta_{\mathbf{q}}\,\mathrm{Re}\,\mathcal{F}_{\mathbf{k},\mathbf{k}',\mathbf{q}}\right] = 1 + \int_{\mathbf{k},\mathbf{k}',\mathbf{q}} 2\zeta_{\mathbf{k}}\zeta_{\mathbf{k}'}\zeta_{\mathbf{q}}\,\mathrm{Re}\,\mathcal{F}_{\mathbf{k},\mathbf{k}',\mathbf{q}} + \dots \qquad \text{(8.3.26)}
$$

We see that we can straightforwardly calculate the correlation functions using a Feynman diagram expansion, with the propagator given by $1/[4A_\zeta(k,t)]$ and a single three-point interaction with coefficient $2\,\mathrm{Re}\,\mathcal{F}_{\mathbf{k},\mathbf{k}',\mathbf{q}}$.

8.3.5 Decoherence

Thus far we have written down an expression (8.3.12) for the wave functional $\Psi\left[\zeta\right](t)$, and hence the wave function is

$$|\Psi(t)\rangle = \int_\zeta \Psi\left[\zeta\right](t)|\zeta\rangle \ . \tag{8.3.27}$$

Using this expression we can compute expectation values by writing them as a path integral which admits a solution using the Feynman diagrams.

This is not, however, all that can be done with the wave function. We have seen in the previous subsection that computing expectation values of the fields alone yields an expression (e.g. Eq. (8.3.19)) that depends only on the wave functional as $\Psi^\dagger\Psi$. Such expectation values depend only on the magnitude of the wave function, not its phase. In addition to expectation values, we can also construct the density operator $\hat{\rho} \equiv |\Psi\rangle\langle\Psi|$, which has complex matrix elements $\rho[\zeta,\zeta'] = \Psi[\zeta]\Psi^\dagger[\zeta']$. In particular, we can factorize Hilbert space by partitioning the wavenumbers, assigning those above a cutoff Λ to the "system" and those below Λ to the "environment,"

$$|\zeta\rangle = |S\rangle|E\rangle, \qquad \mathcal{H}_\zeta = \mathcal{H}_S \otimes \mathcal{H}_E \ , \tag{8.3.28}$$

where

$$|S\rangle = \bigotimes_{|\mathbf{k}|>\Lambda} |\zeta_\mathbf{k}\rangle, \qquad |E\rangle = \bigotimes_{|\mathbf{k}|\leq\Lambda} |\zeta_\mathbf{k}\rangle \ . \tag{8.3.29}$$

We can then write the reduced density matrix of the system

$$\begin{aligned} \rho_S[S,S'] &= \langle S|\hat{\rho}_S|S'\rangle = \langle S|\operatorname{Tr}_E\left(|\Psi\rangle\langle\Psi|\right)|S'\rangle \\ &= \langle S|\int \mathcal{D}E\ \langle E|\ \Psi\rangle\langle\Psi|\ E\rangle|S'\rangle = \int \mathcal{D}E\ \Psi[S,E]\Psi^\dagger[S',E] \end{aligned} \tag{8.3.30}$$

where in the last step we have defined the wave functional $\Psi[S,E]$ as the matrix element between $|S\rangle|E\rangle$ and $|\Psi\rangle$:

$$\Psi[S,E] = \left(\langle S| \otimes \langle E|\right)|\Psi\rangle \ . \tag{8.3.31}$$

Decoherence occurs in the system when interactions between the system and the environment cause the decoherence factor (the ratio of the off-diagonal elements of

ρ_S to the diagonal ones) to become small:

$$D[S, S'] \equiv \frac{|\rho_S[S, S']|}{\sqrt{\rho_S[S, S]\rho_S[S', S']}} \ll 1 \ . \tag{8.3.32}$$

Inserting our expression for Ψ (8.3.12) and noting that the Gaussian part (8.3.6) factors as $\Psi_G[\zeta] = \Psi_G^{(S)}[S](t) \times \Psi_G^{(E)}[E](t)$, the decoherence factor becomes

$$D[S, S'] = \left| \frac{\int \mathcal{D}E \ \Psi[S, E]\Psi^\dagger[S', E]}{\sqrt{\left(\int \mathcal{D}E \ \Psi[S, E]\Psi^\dagger[S, E]\right)\left(\int \mathcal{D}E \ \Psi[S', E]\Psi^\dagger[S', E]\right)}} \right|$$

$$= \left| \frac{\int \mathcal{D}E \ \left|\Psi_G^{(E)}[E]\right|^2 \Psi_{NG}[S, E]\Psi_{NG}^\dagger[S', E]}{\sqrt{\left(\int \mathcal{D}E \ \left|\Psi_G^{(E)}[E]\right|^2 |\Psi_{NG}[S, E]|^2\right)\left(\int \mathcal{D}E \ \left|\Psi_G^{(E)}[E]\right|^2 |\Psi_{NG}[S', E]|^2\right)}} \right| \ . \tag{8.3.33}$$

When the non-Gaussian piece of the wave function is a pure phase, which is the case to lowest order in $\mathrm{Re}\,\mathcal{F}/\mathrm{Im}\,\mathcal{F}$ in Section 8.3.3, both integrals in the denominator integrate to one and the decoherence factor simplifies to

$$D[S, S'] = \left| \int \mathcal{D}E \ \left|\Psi_G^{(E)}[E]\right|^2 \Psi_{NG}[S, E]\Psi_{NG}^\dagger[S', E] \right| \ . \tag{8.3.34}$$

The problem is now reduced to performing the calculation with the previously given forms of Ψ_G and Ψ_{NG}. Ref. [248] carries out this calculation for the case of a single super-Hubble mode, $\mathcal{H}_S = \{\zeta_{\mathbf{q}}, \zeta_{\mathbf{q}}^\dagger = \zeta_{-\mathbf{q}}, q < H\}$. As in Section 8.3.4, Eq. (8.3.34) can be written as an expectation value, this time in the theory of the environment modes, and solved in the deeply super-Hubble limit $|q\tau| \ll 1$ using Feynman diagrams and the cumulant expansion. In our notation, the result is [248]

$$D[\zeta_{\mathbf{q}}, \tilde{\zeta}_{\mathbf{q}}](\tau = -1/aH) = \exp\left[-\frac{1}{288}(\epsilon + \eta)^2|\Delta\bar{\zeta}_{\mathbf{q}}|^2 \left(\frac{aH}{q}\right)^3 + \ldots\right], \tag{8.3.35}$$

where the dots indicate terms higher-order in $\mathcal{F}^2$ and $\Delta\bar{\zeta}_{\mathbf{q}} \equiv \bar{\zeta} - \bar{\zeta}_{\mathbf{q}}'$ is the rescaled dimensionless amplitude of $\zeta_{\mathbf{q}} - \zeta_{\mathbf{q}}'$, defined by $\zeta_{\mathbf{q}} \equiv V^{1/2}q^{-3/2}\pi\sqrt{2}\bar{\zeta}_{\mathbf{q}}$. The barred

quantities have variance

$$\left\langle |\bar{\zeta}_{\mathbf{q}}|^2 \right\rangle = \frac{H^2}{2\epsilon M_p^2} \frac{1}{(2\pi)^2} \equiv \Delta_\zeta^2 \, , \tag{8.3.36}$$

and so $\left\langle |\Delta\bar{\zeta}_{\mathbf{q}}|^2 \right\rangle = 2\Delta_\zeta^2$. The dimensionless decoherence "rate" is then the negative log of the decoherence factor with $|\Delta\bar{\zeta}_{\mathbf{q}}|^2$ set equal to its expectation value,

$$\Gamma_{\text{deco}}(q, a) \approx \left(\frac{\epsilon + \eta}{12}\right)^2 \Delta_\zeta^2 \left(\frac{aH}{q}\right)^3 \quad \text{for } q \ll aH \, . \tag{8.3.37}$$

Decoherence has occurred when this rate, and hence the negative of the exponent in the decoherence factor, becomes large. The rate does not grow large until long after Hubble crossing, at $q = aH$, because of the smallness of the slow-roll parameters and the amplitude of fluctuations (constrained by observations of the CMB [253] to be $\Delta_\zeta^2 \sim 10^{-9}$ at 60 e-folds before the end of inflation). For reasonable values of $(\epsilon + \eta) \sim 10^{-5}$–$10^{-2}$, the modes seen in the CMB would have decohered 10–20 e-folds after Hubble crossing.

In the remainder of this paper, we discuss the implications of this delayed decoherence for eternal inflation. In the next section, we establish that decoherence of long-wavelength modes affects the evolution of short-wavelength modes evolving in the decohered long-wavelength background, and argue that this change in evolution implies the backreaction of the Hubble constant required for eternal inflation. We then turn to discussion of the quantitative differences between the resulting picture and the standard picture of stochastic eternal inflation caused by the delay of decoherence far beyond Hubble crossing.

8.4 Branching and Backreaction

As we have shown, the results of Ref. [248] indicate that decoherence of super-Hubble modes due to gravitational interactions alone is inevitable, though the weakness of these interactions means that the modes typically take several Hubble times after Hubble crossing to decohere. Because modes continually expand across the Hubble radius during inflation, they are also continually decohering, so the overall wave function is itself continually branching; on each branch there is a definite classical value for every mode which has become sufficiently long-wavelength. Since long-wavelength and short-wavelength modes interact gravitationally, we expect the short-

wavelength modes to have a different reaction in different branches to the decohered long-wavelength modes.

In this section we formalize this argument, which we have already made schematically in Section 8.2, by introducing the notion of per-branch observables. We show that short-wavelength modes do indeed evolve differently in different branches of the wave function. We argue that this differing evolution indicates the nature of backreaction away from our perturbative picture; short-wavelength modes evolve as if they experience different values of the Hubble constant in different branches, and there exists a gauge choice on which the effective Hubble constant itself differs from branch to branch.

8.4.1 Observables on branches

In the previous section, we calculated the expectation value of products of fields with respect to the overall state $|\Psi\rangle$. Once decoherence has occurred, however, the evolution in a particular decohered branch is not given by this expectation value, but from the expectation value with respect to the state of that particular branch. As discussed in Section 8.3.1 above, every field configuration $|\zeta\rangle$ is an eigenstate of field value for each individual momentum mode. Since the mode decoheres in the field value basis, we can label individual branches by the field value of the decohered mode[3] in that branch, $\zeta^{\star}_{\mathbf{k}_{\mathrm{dec}}}$. The state $|\Psi\rangle$ can thus be projected onto an individual branch by considering only the field configurations on which the field value of the decohered mode is $\zeta^{\star}_{\mathbf{k}_{\mathrm{dec}}}$, then renormalizing.

More precisely, we define the state $|\zeta^{\star}_{\mathbf{k}}\rangle \in \mathcal{H}_{\zeta,\mathbf{k}}$ as the eigenstate of $\hat{\zeta}_{\mathbf{k}}$ with eigenvalue $\zeta^{\star}_{\mathbf{k}}$, as in Eq. (8.3.3). Then $|\zeta^{\star}_{\mathbf{k}}\rangle\langle\zeta^{\star}_{\mathbf{k}}|$ projects states in the Hilbert space factor $\mathcal{H}_{\zeta,\mathbf{k}}$, and we can define an associated projector on the entire Hilbert space $\mathcal{H}_{\zeta}$ by multiplying this projector by the identity on all other factors,

$$\hat{P}^{\zeta^{\star}_{\mathbf{k}}}_{\mathbf{k}} \equiv \left(|\zeta^{\star}_{\mathbf{k}}\rangle \otimes \bigotimes_{\mathbf{k}'\neq\mathbf{k}} \mathbb{1}_{\mathbf{k}'}\right)\left(\langle\zeta^{\star}_{\mathbf{k}}| \otimes \bigotimes_{\mathbf{k}'\neq\mathbf{k}} \mathbb{1}_{\mathbf{k}'}\right) \,, \tag{8.4.1}$$

---

[3]In fact modes larger than $\mathbf{k}_{\mathrm{dec}}$ have also decohered, so properly speaking we must specify the values of all the decohered modes to uniquely label a branch. We neglect this complication, which can easily be incorporated at the cost of complicating the notation, throughout the section. The final Feynman rules presented in Fig. 8.1, however, take the need to consider each decohered mode into account.

whose action on field configurations defined by Eq. (8.3.2) is simply

$$\hat{P}_{\mathbf{k}}^{\zeta_{\mathbf{k}}^{\star}}|\zeta\rangle = \langle\zeta_{\mathbf{k}}^{\star}|\,\zeta_{\mathbf{k}}\rangle\,|\zeta\rangle = \delta_{\zeta_{\mathbf{k}}^{\star},\zeta_{\mathbf{k}}}|\zeta\rangle \ . \tag{8.4.2}$$

We can now repeat the calculation in Section 8.3.4 above for a branch with a definite field value $\zeta_{\mathbf{k}_{\mathrm{dec}}}^{\star}$ in the $\mathbf{k}_{\mathrm{dec}}$-th mode:

$$
\begin{aligned}
\langle\zeta_{\mathbf{k}_{\star}}\zeta_{-\mathbf{k}_{\star}}\rangle_{\zeta_{\mathbf{k}_{\mathrm{dec}}}^{\star}} &= \frac{1}{N_{\zeta_{\mathbf{k}_{\mathrm{dec}}}^{\star}}}\left\langle\Psi\right|\hat{P}_{\mathbf{k}_{\mathrm{dec}}}^{\zeta_{\mathbf{k}_{\mathrm{dec}}}^{\star}}\hat{\zeta}_{\mathbf{k}_{\star}}\hat{\zeta}_{-\mathbf{k}_{\star}}\hat{P}_{\mathbf{k}_{\mathrm{dec}}}^{\zeta_{\mathbf{k}_{\mathrm{dec}}}^{\star}}\left|\Psi\right\rangle \\
&= \frac{1}{N_{\zeta_{\mathbf{k}_{\mathrm{dec}}}^{\star}}}\left\langle\Psi\right|\left(\int_{\zeta}|\zeta\rangle\langle\zeta|\right)\hat{P}_{\mathbf{k}_{\mathrm{dec}}}^{\zeta_{\mathbf{k}_{\mathrm{dec}}}^{\star}}\hat{\zeta}_{\mathbf{k}_{\star}}\hat{\zeta}_{-\mathbf{k}_{\star}}\hat{P}_{\mathbf{k}_{\mathrm{dec}}}^{\zeta_{\mathbf{k}_{\mathrm{dec}}}^{\star}}\left(\int_{\zeta'}|\zeta'\rangle\langle\zeta'|\right)\left|\Psi\right\rangle \\
&= \frac{1}{N_{\zeta_{\mathbf{k}_{\mathrm{dec}}}^{\star}}}\int_{\zeta}\Psi^{\dagger}[\zeta]\int_{\zeta'}\langle\zeta|\hat{P}_{\mathbf{k}_{\mathrm{dec}}}^{\zeta_{\mathbf{k}_{\mathrm{dec}}}^{\star}}\hat{\zeta}_{\mathbf{k}_{\star}}\hat{\zeta}_{-\mathbf{k}_{\star}}\hat{P}_{\mathbf{k}_{\mathrm{dec}}}^{\zeta_{\mathbf{k}_{\mathrm{dec}}}^{\star}}|\zeta'\rangle\Psi[\zeta'] \\
&= \frac{1}{N_{\zeta_{\mathbf{k}_{\mathrm{dec}}}^{\star}}}\int_{\zeta}\Psi^{\dagger}[\zeta]\int_{\zeta'}\zeta_{\mathbf{k}_{\star}}\zeta'_{-\mathbf{k}_{\star}}\left\langle\zeta_{\mathbf{k}_{\mathrm{dec}}}\middle|\zeta_{\mathbf{k}_{\mathrm{dec}}}^{\star}\right\rangle\left\langle\zeta_{\mathbf{k}_{\mathrm{dec}}}^{\star}\middle|\zeta'_{\mathbf{k}_{\mathrm{dec}}}\right\rangle\langle\zeta|\,\zeta'\rangle\,\Psi[\zeta'] \\
&= \frac{1}{N_{\zeta_{\mathbf{k}_{\mathrm{dec}}}^{\star}}}\int_{\zeta}\left\langle\zeta_{\mathbf{k}_{\mathrm{dec}}}^{\star}\middle|\zeta_{\mathbf{k}_{\mathrm{dec}}}\right\rangle^{2}\Psi^{\dagger}[\zeta]\zeta_{\mathbf{k}_{\star}}\zeta_{-\mathbf{k}_{\star}}\Psi[\zeta] \\
&= \frac{1}{N_{\zeta_{\mathbf{k}_{\mathrm{dec}}}^{\star}}}\int_{\zeta}\left\langle\zeta_{\mathbf{k}_{\mathrm{dec}}}^{\star}\middle|\zeta_{\mathbf{k}_{\mathrm{dec}}}\right\rangle^{2}\zeta_{\mathbf{k}_{\star}}\zeta_{-\mathbf{k}_{\star}}\Psi_{G}^{\star}\Psi_{G}\Psi_{NG}^{\dagger}\Psi_{NG}\ , \tag{8.4.3}
\end{aligned}
$$

where the normalization factor is defined so that the wave function on each branch has unit norm, $\left\langle\Psi\right|\hat{P}_{\mathbf{k}_{\mathrm{dec}}}^{\zeta_{\mathbf{k}_{\mathrm{dec}}}^{\star}}\hat{P}_{\mathbf{k}_{\mathrm{dec}}}^{\zeta_{\mathbf{k}_{\mathrm{dec}}}^{\star}}\left|\Psi\right\rangle/N_{\zeta_{\mathbf{k}_{\mathrm{dec}}}^{\star}} = 1$. Again, Eq. (8.4.3) says that we are supposed to integrate only over the field configurations where the decohered mode has the correct field value, i.e. the ones on the appropriate branch.

8.4.2 Feynman rules on branches

In the pure-phase approximation, the integrals over $\zeta_{\mathbf{k}_{\mathrm{dec}}}$ and $\zeta_{\mathbf{k}_{\star}}$ are independent and the extra term contributes only an overall constant of proportionality that cancels in the normalization. In this approximation, evolution of short-wavelength modes is unaffected by decoherence. A better approximation is to treat $\mathrm{Re}\,\mathcal{F}$ as small compared to $\mathrm{Im}\,\mathcal{F}$, yielding Eq. (8.3.24). Inserting this quantity into the two-point

function for a decohered branch $\langle \zeta_{\mathbf{k}_\star} \zeta_{-\mathbf{k}_\star} \rangle_{\zeta^\star_{\mathbf{k}_{\mathrm{dec}}}}$, Eq. (8.4.3) gives

$$\langle \zeta_{\mathbf{k}_\star} \zeta_{-\mathbf{k}_\star} \rangle_{\zeta^\star_{\mathbf{k}_{\mathrm{dec}}}} = \frac{1}{N_{\zeta^\star_{\mathbf{k}_{\mathrm{dec}}}}} \int_\zeta \left\langle \zeta^\star_{\mathbf{k}_{\mathrm{dec}}} \middle| \zeta_{\mathbf{k}_{\mathrm{dec}}} \right\rangle^2 \zeta_{\mathbf{k}_\star} \zeta_{-\mathbf{k}_\star} \exp\left[- \int_{\mathbf{k}} 2\zeta_{\mathbf{k}} \zeta^\dagger_{\mathbf{k}} \mathrm{Re}\, A_\zeta(k,t) \right]$$
$$\times \exp\left[\int_{\mathbf{k},\mathbf{k}',\mathbf{q}} 2\zeta_{\mathbf{k}} \zeta_{\mathbf{k}'} \zeta_{\mathbf{q}} \mathrm{Re}\, \mathcal{F}_{\mathbf{k},\mathbf{k}',\mathbf{q}} \right] .$$
$$(8.4.4)$$

This expression, combined with the Taylor expansion (8.3.26), allows us to compute correlation functions on decohered branches, but actually writing down the equivalent Feynman rules requires some thought. Ultimately (from the path-integral perspective) we can use Feynman diagrams to compute correlation functions because Taylor expansion lets us write each integral over momentum modes in the form of a polynomial multiplied by a Gaussian in a particular momentum mode, which we can compute using Wick's theorem. Only the integrals for which the polynomial is a nontrivial function of the momentum modes yield nontrivial results; the contribution of every other Gaussian is canceled by the denominator. In terms of Feynman diagrams, these canceled expressions are just the disconnected diagrams. For example, in computing the propagator in Eq. (8.3.22) from Eq. (8.3.21), only the terms in the exponential with $\mathbf{k} = \pm \mathbf{k}_\star$ are important.

We can use Feynman diagrams to compute correlation functions in a particular branch, but we need to carefully take into account the extra factor of $\left\langle \zeta^\star_{\mathbf{k}_{\mathrm{dec}}} \middle| \zeta_{\mathbf{k}_{\mathrm{dec}}} \right\rangle^2$, i.e. we need to restrict the path integral to only span over field configurations with nonzero overlap with the branch. This gives a delta function for each decohered mode. We could impose the delta function separately on each diagram containing decohered modes, but we may also immediately use the delta function to integrate over these modes and simplify the path integral. We integrate each integral over the decohered field mode $\zeta_{\mathbf{k}_{\mathrm{dec}}}$ by localizing to the actual value of the mode on the branch, replacing $\zeta_{\mathbf{k}_{\mathrm{dec}}}$ by $\zeta^\star_{\mathbf{k}_{\mathrm{dec}}}$ wherever it appears.

One replacement is in the $\zeta^2_{\mathbf{k}_{\mathrm{dec}}}$ term that is the coefficient of $\mathrm{Re}\, A_\zeta(k_{\mathrm{dec}}, t)$ in (8.4.4). After we have made this replacement, this term yields a ζ-independent normalization factor which cancels in the numerator and denominator. In terms of Feynman diagrams, the propagator factor for a decohered momentum mode is just 1, which is unsurprising because we have set this mode equal to its classical value in the branch. At this point we can simply integrate out the propagating decohered modes entirely; all interactions involving them will involve the insertion of a classical

$$\underline{\qquad}_{\mathbf{k}} \;=\; \frac{1}{4A_\zeta(\mathbf{k})} \qquad\qquad \times\!-\!-\!-\!-\!-\!-\!-\!-_{\mathbf{k}_{\text{dec}}} \;=\; \zeta^*_{\mathbf{k}_{\text{de}}}$$

$$\underset{\mathbf{k}\qquad\quad\mathbf{p}}{\overset{\mathbf{q}\big|}{\underline{\qquad\quad}}} \;=\; \underset{\mathbf{k}\qquad\quad\mathbf{p}}{\overset{\mathbf{q}\,\vdots^{\times}}{\underline{\qquad\quad}}} \;=\; 2\mathrm{Re}(\mathcal{F}_{\mathbf{k},\mathbf{q},\mathbf{p}})\delta^3(\mathbf{k}+\mathbf{q}+\mathbf{p})$$

$$\langle\zeta_{\mathbf{k}_\star}\zeta_{-\mathbf{k}_\star}\rangle_{\zeta^*_{\mathbf{k}_{\text{de}}}} \;=\; \underline{\qquad}_{\mathbf{k}_\star} \;+\; \int_{|\mathbf{q}|>\mathbf{k}_{\text{de}}} \underset{\mathbf{k}_\star\;\;\overset{\mathbf{q}}{\bigcirc}\;\;\mathbf{k}_\star}{{}_{\mathbf{k}_\star-\mathbf{q}}} \;+\; \sum_i \underset{\mathbf{k}_\star}{\overset{\overset{\times}{\vdots}_i\quad\overset{\times}{\vdots}_i}{\underline{\qquad\qquad}}} \;+\;\dots$$

Figure 8.1. Computation of $\langle\zeta_{\mathbf{k}_\star}\zeta_{-\mathbf{k}_\star}\rangle_{\zeta^*_{\mathbf{k}_{\text{dec}}}}$ using Feynman diagrams. As discussed in Section 8.3.4 above, non-decohered modes have a propagator given by $1/[4A_\zeta(k,t)]$ and a single three-point interaction with coefficient $2\,\mathrm{Re}\,\mathcal{F}_{\mathbf{k},\mathbf{k}',\mathbf{q}}$. A decohered mode field insertion comes with a factor of its field value, $\zeta^*_{\mathbf{k}_{\text{dec}}}$. At leading order the $\zeta_{\mathbf{k}_\star}$ two-point function is corrected by diagrams with two interaction vertices. We split the diagrams into two categories: those where no intermediate momenta are decohered, which we write as a loop correction integrating over momenta greater than $\mathbf{k}_{\text{dec}}$, and those involving decohered momenta, which we represent as a sum over diagrams with two field insertions.

external source.

In addition, we need to replace the decohered field modes which appear in the interaction term. We treat each such mode as a frozen classical source, to be inserted as necessary in the propagator for the dynamical short-wavelength modes, as shown in Fig. 8.1. Our assumption of perturbativity allows us to approximate the interaction term by its Taylor expansion truncated at a given order, yielding a polynomial in $\zeta_{\mathbf{k}}$. The delta function means that we need to replace the polynomial with a piecewise function which substitutes $\zeta^*_{\mathbf{k}_{\text{dec}}}$ for $\zeta_{\mathbf{k}_{\text{dec}}}$ on configurations that overlap with the branch and is zero on all other configurations. Again, this substitution takes place in both the numerator and the denominator (normalization factor). At the level of the first quantum corrections, only the lowest-order term in the denominator (the zero-interaction term, with no factors of $\zeta^*_{\mathbf{k}_{\text{dec}}}$) contributes, so there is a contribution to the path integral with two insertions of the decohered modes[4], proportional to $(\zeta^*_{\mathbf{k}_{\text{dec}}})^2$. In terms of Feynman diagrams, each insertion of an external decohered mode gives a factor of $\zeta^*_{\mathbf{k}_{\text{dec}}}$. As expected, the leading correction to the two-point

___

[4]Because interactions conserve momentum, the term with one insertion does not contribute to $\langle\zeta_{\mathbf{k}_\star}\zeta_{-\mathbf{k}_\star}\rangle_{\zeta^*_{\mathbf{k}_{\text{dec}}}}$, which has equal ingoing and outgoing short-wavelength momentum.

function of a non-decohered field is proportional to the square of the field value of the classical $\zeta_{\mathbf{k}_{\text{dec}}}$ field. This confirms our intuition that short-wavelength modes should evolve differently in different branches.

In summary, the Feynman rules, shown in Fig. 8.1, are the following. For non-decohered fields, the propagator is $1/\left(4A_\zeta(k,t)\right)$. For each decohered field $\zeta_{\mathbf{k}_{\text{dec},i}}$ labeled by i, only modes with the specific decohered field value $\zeta^\star_{\mathbf{k}_{\text{dec},i}}$ contribute on a given branch, and only as external sources. For these modes, field insertions give a factor of $\zeta^\star_{\mathbf{k}_{\text{dec},i}}$. All three-point functions among decohered and non-decohered fields have the same interaction vertex, with coefficient $2\mathrm{Re}\,\mathcal{F}_{\mathbf{k},\mathbf{k}',\mathbf{q}}$.

8.4.3 Cosmological evolution

In the previous subsection we established the intuitive result that short-wavelength modes evolving in a particular branch are affected by long-wavelength modes as if they are evolving in a particular classical background[5], namely the solution to the Einstein equations with the particular nonzero values of the ζ field at long wavelengths (i.e. field values $\zeta^\star_{\mathbf{k}_{\text{dec}}}$) that characterize the branch. In general these geometries, unlike our initial background cosmology, will have nonzero (and nontrivial) spatial curvature. Reproducing the usual eternal inflation story requires transforming to a gauge where the spatial curvature is once again zero, in which we expect that the geometries on various branches of the wave function will have different Hubble constants. This is a standard procedure in the eternal inflation literature (see e.g. Ref. [153]) and we only sketch out the steps schematically.

We first switch from the ζ basis, where the probability distribution over field values is given in the pure phase approximation by Eq. (8.3.6), to the basis of inflaton field values $\phi_{\mathbf{k}}$ in which the eternal inflation picture is usually developed. In the inflaton field gauge, the propagating degree of freedom is the variation $\delta\phi$ of the

---

[5]In single-field slow-roll inflation, the three-point function $\langle \zeta_{\mathbf{q}} \zeta_{\mathbf{k}_{\text{ph}}} \zeta_{\mathbf{k}'_{\text{ph}}} \rangle'$ in "physical coordinates" vanishes in the squeezed limit, $\mathbf{q} \to 0$ [254, 255], where $k_{\text{ph}} \equiv k(1 - \zeta_L)$ and the prime indicates the removal of the momentum-conserving delta function. The vanishing correlation between short-wavelength modes and long-wavelength modes in these coordinates might seem in contradiction with our claim that the evolution of the short-wavelength modes depends on the value of the long-wavelength modes. However, decoherence does not change the value of expectation values with respect to the overall wave function $|\Psi\rangle$. Our claim is that the evolution of short-wavelength modes *on each individual branch* depends on the long-wavelength field values which characterize the branch. As previously discussed, this evolution is distinct from the evolution of short-wavelength modes in the overall wave function. The short-wavelength modes are thus uncorrelated with long-wavelength modes in expectation values with respect to the overall wave function, but not with respect to individual branches.

inflaton field from its expectation value. The power spectrum is that of a light scalar field in de Sitter space:

$$\langle \delta\phi_{\mathbf{k}} \delta\phi_{\mathbf{k'}} \rangle = (2\pi)^3 \, \delta\left(\mathbf{k} + \mathbf{k'}\right) \frac{2\pi^2}{k^3} \left(\frac{H}{2\pi}\right)^2. \tag{8.4.5}$$

Just as the ζ power spectrum defines the coefficient $A_\zeta(k,t)$ of the kinetic term in the action via Eq. (8.3.22)—and hence the wave function through Eq. (8.3.6)—the $\delta\phi$ power spectrum defines a new coefficient $A_{\delta\phi}$. We can therefore rewrite Eq. (8.3.6) in the inflaton field value basis by replacing $A_\zeta(k,t) \to A_{\delta\phi}(k,t)$. This is simply a change of variables which does not alter the wave function itself: we are merely shifting a constant factor $1/2\epsilon$ between the coefficient A and the field variable. In particular, the branching structure of the wave function itself is preserved: decoherence gives definite values of long-wavelength $\delta\phi$ modes just as it gives definite values of long-wavelength ζ modes. For the rest of the paper, it is convenient to work with the resulting distribution of inflaton field values.

On each branch of the wave function, we treat the decohered mode as a delta-function momentum-space perturbation of the inflaton field away from its background value. This perturbation breaks the isotropy of the system, so we can no longer solve for the cosmological evolution using the Friedmann equations, but we can instead use perturbation theory around the initial de Sitter background (e.g. Ref. [152]) to compute the shift in the spatial geometry. Finally, we change gauges to one in which the spatial part of the metric is again homogenous and isotropic. This yields a probability distribution over de Sitter regions with different values of the Hubble parameter H, producing branches on which inflation proceeds at different rates. The usual practice in the eternal inflation literature is to instead say that inflation proceeds at different rates in separate spatial regions in a single overall spacetime. We will comment further on this interpretation in the Discussion below.

8.5 Eternal Inflation

Our goal in this section is to consider how the classical picture of slow-roll inflation, in which the cosmology of a region of space undergoing inflation simply responds to the expectation value of the inflaton field, is modified when we include decoherence and branching. Following the existing literature on eternal inflation and the stochastic approximation, we work directly with Fourier modes of the inflaton field ϕ rather than the adiabatic curvature perturbation ζ. As noted in the previous section, even

though we established decoherence in the ζ field value basis, branches with definite values of $\zeta_{\mathbf{k}}$ should also have definite values of $\phi_{\mathbf{k}}$.

8.5.1 The distribution of branches after decoherence

Although we have seen that modes are continually decohering as they grow larger than the decoherence scale $\mathbf{k}_{\text{dec}}^{-1}$, it suffices to follow the evolution of one particular mode, with expectation value $\phi_\star$ at the time it grows beyond the Hubble radius. First consider the classical evolution. Recall the Friedmann equations:

$$H^2 = \rho/3, \qquad \frac{\ddot{a}}{a} = -\left(\rho + 3p\right)/6 , \qquad (8.5.1)$$

where as in Section 8.2 we have set $8\pi G = c = 1$. A scalar field obeys the Klein-Gordon equation,

$$\ddot{\phi} + 3H\dot{\phi} = -V' , \qquad (8.5.2)$$

where $' = d/d\phi$, and has energy density

$$\rho = \dot{\phi}^2/2 + V(\phi) . \qquad (8.5.3)$$

In the slow-roll regime, $\ddot{\phi} \ll 3H\dot{\phi}, -V'$ and $\dot{\phi}^2 \ll V$, and the field value evolves classically at a rate

$$\dot{\phi} = -\frac{V'}{3H} . \qquad (8.5.4)$$

In one Hubble time the classical change is therefore

$$\Delta\phi_{\text{c}} \equiv \dot{\phi}H^{-1} = -\frac{V'}{3H^2} . \qquad (8.5.5)$$

Meanwhile, the dispersion around the classical value [154, 226, 244, 245, 249] obeys Eq. (8.2.2), so the variance accumulated in a single Hubble time is

$$\Delta_{\text{q}}^2 \equiv \left(\langle\delta\phi^2(t + \Delta t)\rangle - \langle\delta\phi^2(t)\rangle\right)_{\Delta t = H^{-1}} = \frac{H^2}{4\pi^2} . \qquad (8.5.6)$$

The overall variance of $\delta\phi$ continues to grow as modes expand past Hubble crossing, but the variance of individual modes freezes out once they exceed the Hubble scale, with variance Δ_{q}^2.

We are interested in what happens after N e-folds after Hubble crossing, where N is the number of e-folds at which modes decohere, which we write explicitly for
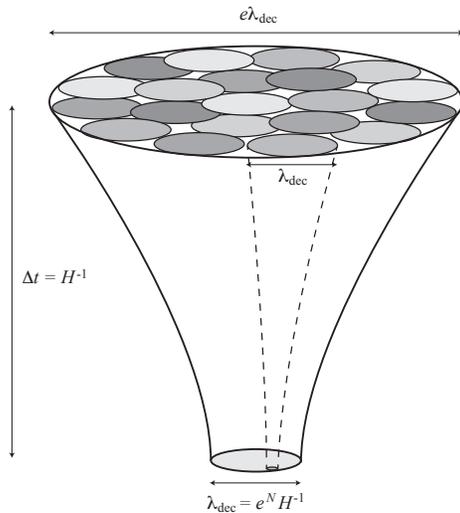
Figure 8.2. The evolution of patches in eternal inflation. We choose to look at an initial patch of linear size given by the wavelength at which modes decohere, N e-folds after Hubble crossing, $\lambda_{\mathrm{dec}} = e^N H^{-1}$. One Hubble time later, the linear size of this comoving region has expanded by e, so the volume now contains $e^3 \approx 20$ patches of the size of the original region.

a general slow-roll potential $V(\phi)$ below. At this time the particular mode we are following, now with size $\lambda_{\mathrm{dec}} \equiv e^N H^{-1}$, decoheres into branches. On each branch of the wave function, the mode has a definite classical value, and the probability distribution of these classical values is given by a Gaussian with width Δ_{q} and mean $\phi_\star + N\Delta\phi_{\mathrm{c}}$:

$$P(\phi) \equiv \frac{1}{\sqrt{2\pi\Delta_{\mathrm{q}}^2}} \exp\left[-\frac{(\phi - \phi_\star - N\Delta\phi_{\mathrm{c}})^2}{2\Delta_{\mathrm{q}}^2}\right] , \tag{8.5.7}$$

where the prefactor ensures proper normalization of the probability distribution.

Note that V' and H are both properly functions of ϕ, so the classical change $\Delta\phi_{\mathrm{c}}$ also depends on the inflaton's location on the potential. In Eq. (8.5.7) we have neglected this effect and assumed that $\Delta\phi_{\mathrm{c}}$ is constant over the range of field values we are interested in, so that the total classical rolling over N e-folds is just $N\Delta\phi_{\mathrm{c}}$. We will relax this assumption below when we consider corrections to the standard eternal inflation picture.

8.5.2 The regime of eternal inflation

Eq. (8.5.7) gives the probability distribution over field values for decohered inflaton modes. Given this probability distribution, when does eternal inflation occur? We are concerned with computing the change in eternal inflation due to delayed decoherence, so we first give the conventional account of eternal inflation [226, 228, 244–246]. We need to compare the expectation value $\langle \phi(t = t_0) \rangle$ of the mode of interest at some initial time t_0 before decoherence has occurred to its value in particular decohered branches, drawn from the probability distribution $P(\phi)$, which is defined at the time of decoherence, $t = t_0 + \Delta t$. The probability that the field on a particular branch has moved up its potential is given by

$$\mathrm{Pr}(\phi > \langle \phi(t = t_0) \rangle) \equiv \int_{\langle \phi(t=t_0) \rangle}^{\infty} P(\phi) d\phi \ . \tag{8.5.8}$$

Because $P(\phi)$ is supported on all values of ϕ, the probability that the field on a particular branch has moved up its potential is always strictly nonzero. When the probability is large enough, however, we say that the entire ensemble of branches, i.e. the wave function, is undergoing eternal inflation. Here "large enough" is usually taken to mean larger than the reciprocal of the growth in volume during this time: $\mathrm{Pr}(\phi > \langle \phi_0 \rangle) \gtrsim e^{-3H\Delta t}$.

This criterion for eternal inflation to occur is usually justified in terms of the growth of the volume of inflating spacetime. The situation is depicted in Fig. 8.2. Consider a volume of space with initial size given by the decoherence length $\lambda_{\mathrm{dec}} \equiv e^N H^{-1}$. In the time Δt it takes for a given mode to reach the scale λ_{dec} and decohere, the initial volume will have grown by a factor $e^{3H\Delta t}$. We can therefore divide the volume into $e^{3H\Delta t}$ regions with volume equivalent to the initial one. We imagine for now that decoherence results in a separate classical field value in each of these regions (we will discuss the validity of this assumption later). Hence if the probability of moving up the potential in a given region is larger than $e^{-3H\Delta t}$, a typical branch of the wave function describing the evolution of the entire initial volume will contain at least one region of the same size as that initial volume where the field has moved up on the potential and the rate of expansion has increased. In this case inflation is said to be "self-reproducing" or eternal. It remains only to choose a convenient timescale. The physically relevant timescale in the problem is the Hubble time H^{-1}, which leads to the familiar criterion that eternal inflation occurs if there is a probability to move up the potential of at least $e^{-3} \approx 5\%$.

Accordingly, consider the situation one Hubble time before decoherence occurs.

Subject to the assumptions discussed at the end of Subsection 8.5.1, the expectation value of the mode of interest is then

$$\langle \phi(t = t_0) \rangle = \phi_\star + (N - 1)\Delta\phi_c = \phi_\star + (N - 1)V'/3H^2 , \qquad (8.5.9)$$

where again $\phi_\star$ is the field value at Hubble crossing, while the variance, which has been frozen out since Hubble crossing, remains $\Delta_q^2 = H^2/4\pi^2$. Now wait for one last Hubble time. The volume of the inflating space expands by a factor of $e^3 \approx 20$, and the expectation value of the field changes to $\phi_\star + N\Delta\phi_c$.

The probability that the field has effectively "jumped" up the potential compared to where it was an e-fold ago is given by the proportion of the probability distribution where $\phi > \phi_\star + (N - 1)\Delta\phi_c$:

$$\Pr\left(\phi > \phi_\star + (N - 1)\Delta\phi_c\right) \equiv \int_{\phi_\star + (N-1)\Delta\phi_c}^{\infty} P(\phi)d\phi = \frac{1}{2}\left[1 - \mathrm{erf}\left(\frac{-\Delta\phi_c}{\Delta_q\sqrt{2}}\right)\right] .$$
$$(8.5.10)$$

Recall that the error function $\mathrm{erf}(x)$ ranges from 0 to 1 as x ranges from 0 to ∞. So a large probability of jumping up the potential requires that the quantum dispersion is large compared to the classical rolling.

Notice that the final expression in Eq. (8.5.10) lacks any direct dependence on N, the number of e-folds from Hubble crossing to decoherence. Hence when the expression is valid we recover exactly the standard predictions of eternal inflation.

We can now insert the details of the inflationary potential. First, the argument of the error function is

$$\frac{-\Delta\phi_c}{\Delta_q\sqrt{2}} = \frac{\pi\sqrt{2}V'}{3H^3} = \frac{2\pi\sqrt{\epsilon}}{H} , \qquad (8.5.11)$$

where we have used $\epsilon = (V'/V)^2/2$, $H^2 = V/3$. Slow-roll eternal inflation in the sense we have described above occurs when

$$\Pr\left[\phi > \phi_\star + (N - 1)\Delta\phi_c\right] > e^{-3} . \qquad (8.5.12)$$

Eqs. (8.5.10) and (8.5.11) let us check where this is true for a given potential given the Hubble parameter H and slow-roll parameters ϵ and η. We see from Eq. (8.5.11) that quantum fluctuations become more important for flatter potentials (small ϵ) and at greater energy scales (large H/M_p).

8.5.3   Corrections from delayed decoherence

In deriving Eq. (8.5.10) we assumed, as discussed at the end of Subsection 8.5.1, that the rate of classical rolling $\Delta\phi_{\rm c}$ was constant over the range of e-folds from Hubble crossing to decoherence and hence that the total classical rolling in this time was just $N\Delta\phi_{\rm c}$. In this subsection we investigate the slight corrections which result from relaxing this assumption. We focus on determining the range of ϕ values in which modes that cross the Hubble scale freeze out with sufficiently large variance to allow for eternal inflation.

As explained in the last subsection, we are interested in the last e-fold of classical expansion before decoherence occurs. Denote the value of ϕ at the start of this interval by ϕ_s and at the end by ϕ_e. As above, the value of ϕ when the mode of interest crossed the Hubble scale is denoted by $\phi_\star$. We can now rewrite the probability distribution of classical field values at decoherence as

$$P(\phi) \equiv \frac{1}{\sqrt{2\pi\Delta_{\rm q}^2\left(\phi_\star\right)}} \exp\left[-\frac{\left(\phi - \phi_e\right)^2}{2\Delta_{\rm q}^2}\right] \tag{8.5.13}$$

and the probability of moving upward on the potential as

$$\Pr\left(\phi > \phi_s\right) \equiv \int_{\phi_s}^{\infty} P(\phi)d\phi = \frac{1}{2}\left[1 - \mathrm{erf}\left(\frac{-\left(\phi_s - \phi_e\right)}{\Delta_{\rm q}\left(\phi_\star\right)\sqrt{2}}\right)\right] . \tag{8.5.14}$$

If the field is still in the slow-roll regime at the time that the mode of interest decoheres, Eq. (8.5.5) is still valid:

$$\phi_s - \phi_e \approx \dot\phi H^{-1} = -\frac{V'}{3H^2}, \tag{8.5.15}$$

but now we should evaluate V' and H during the last e-fold of inflation before decoherence, say at $\left(\phi_s + \phi_e\right)/2$, rather than at Hubble crossing.

We would like to evaluate Eq. (8.5.15) and thus Eq. (8.5.14) as a function of the field value at horizon crossing, $\phi_\star$. A first approximation is to take

$$\phi_s - \phi_e \approx -\left.\frac{V'}{3H^2}\right|_{\phi=\phi_\star}, \tag{8.5.16}$$

but this simply reproduces the N-independent expression for $\Pr(\phi)$ given in the previous expression. If we are far enough in the slow-roll regime, $N\ddot\phi \ll 3H\dot\phi$, we

can do better by evaluating H and ϕ at the first-order approximation to $(\phi_s + \phi_e)/2$, i.e. $\phi_\star + (N - 1/2)\Delta\phi_c$:

$$\phi_s - \phi_e \approx - \left.\frac{V'}{3H^2}\right|_{\phi=\phi_\star-\left(N-\frac{1}{2}\right)\frac{V'}{3H^2}\big|_{\phi=\phi_\star}} . \qquad (8.5.17)$$

This expression may then straightforwardly be evaluated for a given potential. Notably, a dependence on N has now been reintroduced. Using Eqs. (8.3.37) and (8.3.36),

$$N \equiv \left(\ln\frac{aH}{q} \quad \text{s.t. } \Gamma_{\text{deco}} = 1\right) = -\frac{1}{3}\ln\frac{H^2\left(\epsilon+\eta\right)^2}{1152\pi^2\epsilon} \approx 3.11 - \frac{1}{3}\ln\frac{H^2\left(\epsilon+\eta\right)^2}{\epsilon} . \qquad (8.5.18)$$

At the order we are working it is consistent to evaluate this expression at $\phi = \phi_\star$.

As a worked example, Figure 8.3 plots the two expressions (8.5.10) and (8.5.14) for a ϕ^4 potential. For this potential $N(\phi)$ decreases logarithmically with ϕ, from 9.38 at $\phi = 100$ to 7.85 at $\phi = 1000$. This delayed decoherence has only a small effect on the probability of eternal inflation, changing the probability by order 10^{-5}.

8.6 Discussion

In the previous section we have largely worked within the standard picture of eternal inflation, altering it only by changing when the onset of decoherence occurs. In the process we have noted a few uncertainties regarding this picture, which to our knowledge have not been fully resolved.

One ambiguity is the value of Δt, the time interval at which we calculate how the wave function has branched (or in conventional language, at which quantum jumps occur). Equivalently, this is the time before decoherence at which we take the expectation value $\langle\phi\rangle$, in order to compare it to the distribution $P(\phi)$ of values of the field in decohered branches, and therefore evaluate the probability that the field has jumped up in its potential, allowing for eternal inflation. We have chosen $\Delta t = H^{-1}$, which reproduces the criterion that inflation is eternal when at least 5% of patches have jumped upward on the potential. Note that this implies that $N = 1$ in the standard picture, which corresponds to decoherence occurring one e-fold after Hubble crossing, *not* at Hubble crossing itself—a fact which does not seem to be commonly appreciated but is implicit in early work on eternal inflation such as Ref. [154]. The criterion for when eternal inflation occurs depends on Δt, though
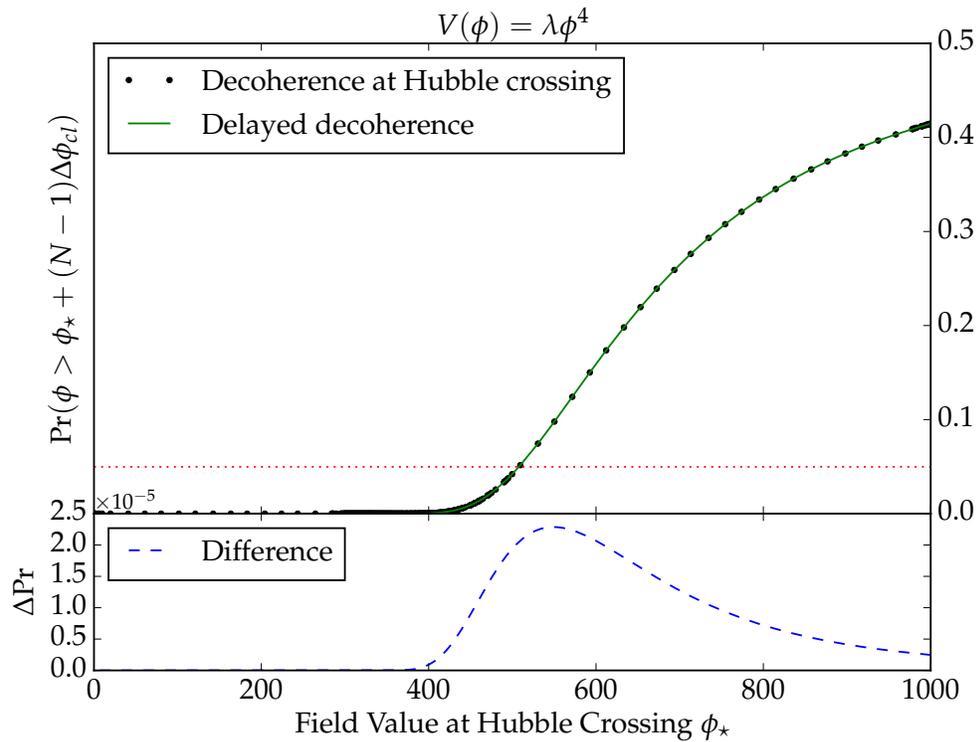
Figure 8.3. Eternal inflation for a ϕ^4 potential. We have set $\lambda \approx 4.28 \times 10^{-14}$, which is the value required to reproduce the amplitude of fluctuations in the CMB: $\Delta_\zeta^2 \approx 2.5 \times 10^{-9}$ 60 e-folds before the end of inflation. On the top plot, the green solid line plots the probability of eternal inflation for modes passing the Hubble scale at a field value $\phi_\star$ using Eqs. (8.5.14) and (8.5.17); the black dots show the result using Eq. (8.5.10). The red dotted horizontal line shows the probability value required for eternal inflation, $e^{-3} \approx 0.05$. The bottom plot shows the difference between the two expressions: the difference in probabilities has a value of around 10^{-5} at field values $\phi_\star \sim 500$ near the lower end of the regime where eternal inflation is allowed. The difference in probabilities is always positive because $\lambda\phi^4$ is concave up, so moving downward on the potential decreases V' and thus the classical rolling per e-fold.

only slightly, since it changes the field value at which we should evaluate the classical rolling.

We are therefore left with the perhaps disquieting fact that whether or not inflation is eternal does not seem to be entirely objective, but rather depends on our choice of discretization. For now, we note that two alternate choices of Δt seem

unsatisfactory. Comparing the situation at decoherence to the situation at Hubble crossing itself, $\Delta t = NH^{-1}$, neglects the fact that in this time many other modes have decohered, making eternal inflation seem harder to achieve than it should actually be. On the other hand, making the approximation that decoherence is instantaneous, $\Delta t = 0$, in addition to being physically unrealistic, simply gives a probability of 50% that the field value increased, which does not seem to match our intuition that eternal inflation should depend on the details of the inflaton potential. So for the moment our choice of $\Delta t = H^{-1}$ seems most natural, in addition to most directly allowing for comparison to the standard picture. We hope to return to this issue in future work. One possibility is that, instead of assuming that decoherence happens immediately, we should be more careful in computing the timescale over which decoherence occurs and inserting this timescale in our calculations. Another possibility, as we now discuss, is that the comparison of field values before and after decoherence is not the appropriate way to determine whether inflation is eternal.

A second, perhaps more serious, issue is the tension between a traditional semi-classical spacetime picture, in which branches of the wave function represent particular spacetimes in which the inflaton takes on slightly different values in nearby patches of space, versus a more intrinsically quantum picture, in which the wave function itself is primary and spacetime is emergent. Establishing that decoherence has occurred means that we can write the wave function in terms of non-interfering branches, each of which has a definite classical value of the decohered mode. It is not clear how we should take into account different probabilities for our universe to emerge from reheating in each of these branches (though one of us has considered a more general version of this question [256]), and/or whether we should consider the different rates of expansion in the different branches. This question seems intimately related to the inflationary measure problem (for reviews, see, e.g., [257, 258]). Some authors have argued that there is a coherent picture of different inflating regions as present in a single spacetime [208], others that the multiverse must be thought of as inherently quantum [259]. We hope to consider this question more extensively in future work. One step in this direction might include more fully carrying out the program sketched in Section 8.4.3 to explicitly derive the wave function of an inflating scalar field in terms of branches with definite values of the Hubble parameter.

8.7 Conclusion

In this paper we have tried to place the assumptions of decoherence and backreaction required for slow-roll eternal inflation on a firmer quantum-mechanical footing. In single-field slow-roll inflation, we can definitively establish the decoherence properties of the inflaton by considering spatial perturbations around a background de Sitter metric. In this gauge the leading interaction is a gravitationally sourced cubic one (8.3.18) whose strength depends on the parameters of the inflaton potential, so that in the slow-roll regime inflaton modes do not typically decohere until they have become very long-wavelength, several e-folds after they pass the Hubble scale (8.5.18). When decoherence has occurred, we have shown that the evolution of inflaton modes is different on different decohered branches of the wave function, each representing a different classical spacetime. Hence the daughter cosmologies after decoherence has occurred have the differing cosmological evolutions required for the eternal inflation mechanism. We can use this backreaction to reproduce the standard predictions for the regime of eternal inflation given a potential, and compute the (typically small) numerical changes to the boundaries of this regime.

Acknowledgements

8.A Free Hamiltonian and Green Function

In this Appendix we derive the free Hamiltonian in Eq. (8.3.10) and the Green function in Eq. (8.3.11) in the Schrödinger picture. We begin with the quadratic

action for ζ (8.3.9), setting $c_s = 1$. To first order[6], the conjugate momentum of ζ is

$$\pi^{(\zeta)} = \frac{\partial \mathcal{L}}{\partial \dot{\zeta}} = 2\epsilon M_p^2 a^3 \dot{\zeta} , \qquad (8.A.1)$$

which obeys the canonical commutation relation $[\zeta(\mathbf{x}), \pi^{(\zeta)}(\mathbf{y})] \equiv i\delta^3(\mathbf{x} - \mathbf{y})$. Although we will write quantities as function of τ, recall that we defined the overdot notation to denote derivatives with respect to t. We use the Fourier transform $\zeta_{\mathbf{k}} = \int d^3\mathbf{x}\zeta(\mathbf{x})e^{-i\mathbf{k}\cdot\mathbf{x}}$ to write the conjugate momentum in terms of its wavelength modes

$$\pi_{\mathbf{k}}^{(\zeta)} = 2\epsilon M_p^2 a^3 \dot{\zeta}_{\mathbf{k}} , \qquad (8.A.2)$$

which are still functions of time. Hence the free Hamiltonian is

$$\hat{H}_{\text{free}}[\zeta] = \int d^3\mathbf{x} \left[\pi^{(\zeta)}\dot{\zeta} - \mathcal{L} \right] = (2\epsilon M_p^2 a^3) \int d^3\mathbf{x} \left[\dot{\zeta}^2 - \frac{1}{2}\left(\dot{\zeta}^2 - \frac{1}{a^2}(\partial_i\zeta)^2 \right) \right]$$
$$= \frac{1}{2} \int_{\mathbf{k}} \left[\frac{1}{2\epsilon M_p^2 a^3} \pi_{\mathbf{k}}^{(\zeta)} \pi_{-\mathbf{k}}^{(\zeta)} + 2\epsilon M_p^2 a k^2 \zeta_{\mathbf{k}}\zeta_{-\mathbf{k}} \right] , \qquad (8.A.3)$$

which matches Eq. (8.3.10). For convenience, we define

$$\int_{\mathbf{k}} \equiv \int \frac{d^3\mathbf{k}}{(2\pi)^3} \qquad \text{and} \qquad \int_{\mathbf{k},\mathbf{k}',\mathbf{q}} \equiv \int \frac{d^3\mathbf{k}}{(2\pi)^3}\frac{d^3\mathbf{k}'}{(2\pi)^3}\frac{d^3\mathbf{q}}{(2\pi)^3} (2\pi)^3 \delta^3(\mathbf{k} + \mathbf{k}' + \mathbf{q}) .$$
$$(8.A.4)$$

With this Hamiltonian and the assumed form of the wave function in Eq. (8.3.6), we expand both sides of the free Schrödinger equation (8.3.5)

$$i\frac{d}{dt}\Psi_G[\zeta](\tau) = \hat{H}_{\text{free}}[\zeta]\Psi_G[\zeta](\tau) . \qquad (8.A.5)$$

For the left-hand side of this equation, we find

$$i\frac{d}{dt}\Psi_G^{(\zeta)}[\zeta](\tau) = i\Psi_G^{(\zeta)}[\zeta](\tau) \left(\frac{\dot{N}_\zeta}{N_\zeta} - \int_{\mathbf{k}} \zeta_{\mathbf{k}}\zeta_{-\mathbf{k}}\dot{A}_\zeta(k,\tau) \right) . \qquad (8.A.6)$$

For the right-hand side, we must act with the conjugate momentum on the wave

---

[6]It suffices to work at lowest order because the terms generated by quadratic corrections cancel in the Hamiltonian density up to cubic order; see footnote 18 of Ref. [248].

function, and thus we express it as a functional derivative: $\pi_{\mathbf{k}}^{(\zeta)} = -i\delta/\delta\zeta_{-\mathbf{k}}$. We find

$$\pi_{\mathbf{k}}^{(\zeta)}\Psi_G^{(\zeta)}[\zeta](\tau) = i\zeta_{\mathbf{k}}\left[A_\zeta(-k,\tau) + A_\zeta(k,\tau)\right]\Psi_G^{(\zeta)}[\zeta](\tau) \tag{8.A.7}$$

$$\pi_{-\mathbf{k}}^{(\zeta)}\Psi_G^{(\zeta)}[\zeta](\tau) = i\zeta_{-\mathbf{k}}\left[A_\zeta(-k,\tau) + A_\zeta(k,\tau)\right]\Psi_G^{(\zeta)}[\zeta](\tau) \tag{8.A.8}$$

$$\pi_{\mathbf{k}}^{(\zeta)}\pi_{-\mathbf{k}}^{(\zeta)}\Psi_G^{(\zeta)}[\zeta](\tau) = (2\pi)^3\left[A_\zeta(-k,\tau) + A_\zeta(k,\tau)\right]\Psi_G^{(\zeta)}[\zeta](\tau)$$
$$- \zeta_{\mathbf{k}}\zeta_{-\mathbf{k}}\left[A_\zeta(-k,\tau) + A_\zeta(k,\tau)\right]^2\Psi_G^{(\zeta)}[\zeta](\tau). \tag{8.A.9}$$

The right-hand side of the free Schrödinger equation becomes

$$\hat{H}_{\text{free}}(t)\Psi_G^{(\zeta)}[\zeta](\tau) = \frac{1}{2}\int_{\mathbf{k}}\left[(2\pi)^3 f_\zeta 2A(k,\tau)\right.$$
$$\left. - f_\zeta(2A(k,\tau))^2\zeta_{\mathbf{k}}\zeta_{-\mathbf{k}} + \frac{1}{f_\zeta}\frac{k^2}{a^2}\zeta_{\mathbf{k}}\zeta_{-\mathbf{k}}\right]\Psi_G^{(\zeta)}[\zeta](\tau) , \tag{8.A.10}$$

where

$$f_\zeta(\tau) \equiv \frac{1}{2\epsilon M_p^2 a^3} = -\frac{\tau^3 H^3}{2\epsilon M_p^2} . \tag{8.A.11}$$

We are interested in solving for A, so we match the terms proportional to $\zeta_{\mathbf{k}}\zeta_{-\mathbf{k}}$ to obtain the differential equation

$$\dot{A} = -2if_\zeta A^2 + \frac{i}{2f_\zeta}\frac{k^2}{a^2} . \tag{8.A.12}$$

After making a change of variables to $a = \exp(Ht)$ and defining

$$A = \frac{aH}{2if_\zeta(a)}\frac{du}{da}\frac{1}{u} , \tag{8.A.13}$$

the differential equation becomes [242]

$$a^2\frac{d^2u}{da^2} + 4a\frac{du}{da} + \frac{k^2}{H^2a^2}u = 0 . \tag{8.A.14}$$

This is the Klein-Gordon equation in de Sitter, which can be solved in terms of Bessel functions. We define $u = x^{3/2}y$ and change variables to $x = k/aH = -k\tau$ to obtain

$$x^2\frac{d^2y}{dx^2} + x\frac{dy}{dx} + \left(x^2 - \nu^2\right)y = 0 , \tag{8.A.15}$$

where $\nu = 3/2$, and the solutions are the Bessel functions of the first and second kinds. To find the correct form of $y(x)$, we need to apply initial conditions: we assume that in the far past ($a \to 0$ or $x \to \infty$ or $\tau \to -\infty$ or $t \to -\infty$) space is de Sitter and thus the solution is quasistatic, $dA/dt|_{t \to -\infty} = 0$. The limiting form of y becomes

$$y \to u_0 x^{-3/2} e^{-ix} . \tag{8.A.16}$$

The appropriate combination of Bessel functions that give the $\exp(-ix)$ dependence is the Hankel function of the 2nd kind, $H_\nu^{(2)}(x)$. For $\nu = 3/2$,

$$y(x) = H_{3/2}(x) = -\sqrt{\frac{2}{\pi x}} \left(1 - \frac{i}{x} \right) e^{-ix} . \tag{8.A.17}$$

Substituting y for A, we find

$$A_\zeta(k, \tau) = k^3 \frac{\epsilon M_p^2}{H^2} \frac{1 - \frac{i}{k\tau}}{1 + k^2 \tau^2} , \tag{8.A.18}$$

which is our desired result. Note that this expression differs by a factor of 2 from Eq. (5.4) of Ref. [248].

Bibliography

[1] J. S. Cotler, G. R. Penington, and D. H. Ranard, "Locality from the Spectrum," arXiv:1702.06142 [quant-ph].

[2] S. Carroll and A. Singh, "To appear,".

[3] H. D. Zeh, "On the interpretation of measurement in quantum theory," *Foundations of Physics* **1** (Mar., 1970) 69–76.

[4] W. Zurek, "Pointer Basis of Quantum Apparatus: Into What Mixture Does the Wave Packet Collapse?," *Phys. Rev.* **D24** (1981) 1516.

[5] R. B. Griffiths, "Consistent histories and the interpretation of quantum mechanics," *J. Statist. Phys.* **36** (1984) 219.

[6] E. Joos and H. D. Zeh, "The Emergence of classical properties through interaction with the environment," *Z. Phys.* **B59** (1985) 223–243.

[7] M. Schlosshauer, "Decoherence, the measurement problem, and interpretations of quantum mechanics," *Rev. Mod. Phys.* **76** (2004) 1267, arXiv:quant-ph/0312059 [quant-ph].

[8] R. M. Wald, "Asymptotic behavior of homogeneous cosmological models in the presence of a positive cosmological constant," *Phys. Rev.* **D28** (1983) 2118.

[9] S. Hollands, "Correlators, Feynman diagrams, and quantum no-hair in deSitter spacetime," *Commun. Math. Phys.* **319** (2013) 1, arXiv:1010.5367 [gr-qc].

[10] D. Marolf and I. A. Morrison, "The IR stability of de Sitter QFT: results at all orders," *Phys. Rev.* **D84** (2011) 044040, arXiv:1010.5327 [gr-qc].

[11] S. M. Carroll and A. Chatwin-Davies, "Cosmic Equilibration: A Holographic No-Hair Theorem from the Generalized Second Law," arXiv:1703.09241 [hep-th].

[12] J. M. Maldacena, "The large-N limit of superconformal field theories and supergravity," *Int.J.Theor.Phys.* **38** (1999) 1113–1133, arXiv:hep-th/9711200 [hep-th].

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

[14] O. Aharony, S. S. Gubser, J. M. Maldacena, H. Ooguri, and Y. Oz, "Large-N field theories, string theory and gravity," *Phys.Rept.* **323** (2000) 183–386, arXiv:hep-th/9905111 [hep-th].

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

[16] J. Maldacena and L. Susskind, "Cool horizons for entangled black holes," *Fortsch.Phys.* **61** (2013) 781–811, arXiv:1306.0533 [hep-th].

[17] L. Susskind, "The world as a hologram," *J.Math.Phys.* **36** (1995) 6377–6396, arXiv:hep-th/9409089 [hep-th].

[18] G. 't Hooft, "Dimensional reduction in quantum gravity," arXiv:gr-qc/9310026 [gr-qc].

[19] J. M. Maldacena, "The Large N limit of superconformal field theories and supergravity," *Int.J.Theor.Phys.* **38** (1999) 1113–1133, arXiv:hep-th/9711200 [hep-th].

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

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

[22] A. Lewkowycz and J. Maldacena, "Generalized gravitational entropy," *JHEP* **1308** (2013) 090, arXiv:1304.4926 [hep-th].

[23] R. Orús, "A practical introduction to tensor networks: Matrix product states and projected entangled pair states," *Ann.Phys.* **349** (2014) 117–158, arXiv:1306.2164 [cond-mat.str-el].

[24] G. Vidal, "Class of Quantum Many-Body States That Can Be Efficiently Simulated," *Phys.Rev.Lett.* **101** (2008) 110501, quant-ph/0610099.

[25] B. Swingle, "Entanglement renormalization and holography," *Phys.Rev.* **D86** (2012) 065007, arXiv:0905.1317 [cond-mat.str-el].

[26] X.-L. Qi, "Exact holographic mapping and emergent space-time geometry," arXiv:1309.6282 [hep-th].

[27] F. Pastawski, B. Yoshida, D. Harlow, and J. Preskill, "Holographic quantum error-correcting codes: Toy models for the bulk/boundary correspondence," arXiv:1503.06237 [hep-th].

[28] B. Czech, L. Lamprou, S. McCandlish, and J. Sully, "Integral Geometry and Holography," arXiv:1505.05515 [hep-th].

[29] C. Bény, "Causal structure of the entanglement renormalization ansatz," *New J.Phys.* **15** (2013) 023020, arXiv:1110.4872 [quant-ph]. http://stacks.iop.org/1367-2630/15/i=2/a=023020?key=crossref.9ac05e6502b99b6ba50913ca6cf2d17e.

[30] G. Evenbly and G. Vidal, "Algorithms for entanglement renormalization," *Phys.Rev.* **B79** (2009) 144108, arXiv:0707.1454 [cond-mat.str-el].

[31] R. N. C. Pfeifer, G. Evenbly, and G. Vidal, "Entanglement renormalization, scale invariance, and quantum criticality," *Phys.Rev.* **A79** (2009) 040301, arXiv:0810.0580 [cond-mat.str-el].

[32] G. Evenbly, P. Corboz, and G. Vidal, "Nonlocal scaling operators with entanglement renormalization," *Phys.Rev.* no. 13, (Oct., 2010) 132411, arXiv:0912.2166 [cond-mat.str-el].

[33] Y.-Y. Shi, L.-M. Duan, and G. Vidal, "Classical simulation of quantum many-body systems with a tree tensor network," *Phys.Rev.* **A74** (2006) 022320, quant-ph/0511070.

[34] P. Pfeuty, "The one-dimensional ising model with a transverse field," *Ann.Phys.* **57** (1970) 79 – 90. http://www.sciencedirect.com/science/article/pii/0003491670902708.

[35] T. Hartman and J. Maldacena, "Time Evolution of Entanglement Entropy from Black Hole Interiors," *JHEP* **1305** (2013) 014, arXiv:1303.1080 [hep-th].

[36] B. Swingle, "Constructing holographic spacetimes using entanglement renormalization," arXiv:1209.3304 [hep-th].

[37] G. Evenbly and G. Vidal, "Scaling of entanglement entropy in the (branching) multiscale entanglement renormalization ansatz," *Phys.Rev.* **B89** (2014) 235113, arXiv:1310.8372 [quant-ph].

[38] J. Haegeman, T. J. Osborne, H. Verschelde, and F. Verstraete, "Entanglement Renormalization for Quantum Fields in Real Space," *Phys.Rev.Lett.* **110** (2013) 100402, arXiv:1102.5524 [hep-th].

[39] M. Nozaki, S. Ryu, and T. Takayanagi, "Holographic Geometry of Entanglement Renormalization in Quantum Field Theories," *JHEP* **1210** (2012) 193, arXiv:1208.3469 [hep-th].

[40] A. Mollabashi, M. Nozaki, S. Ryu, and T. Takayanagi, "Holographic Geometry of cMERA for Quantum Quenches and Finite Temperature," *JHEP* **1403** (2014) 098, arXiv:1311.6095 [hep-th].

[41] C. Holzhey, F. Larsen, and F. Wilczek, "Geometric and renormalized entropy in

conformal field theory," *Nucl.Phys.* **B424** (1994) 443–467, arXiv:hep-th/9403108 [hep-th].

[42] P. Calabrese and J. L. Cardy, "Entanglement entropy and quantum field theory," *J.Stat.Mech.* **0406** (2004) P06002, arXiv:hep-th/0405152 [hep-th].

[43] J. Brown and M. Henneaux, "Central charges in the canonical realization of asymptotic symmetries: An example from three dimensional gravity," *Commun.Math.Phys.* **104** (1986) 207–226. http://dx.doi.org/10.1007/BF01211590.

[44] S. Ryu and T. Takayanagi, "Aspects of Holographic Entanglement Entropy," *JHEP* **0608** (2006) 045, arXiv:hep-th/0605073 [hep-th].

[45] T. Jacobson, "Thermodynamics of space-time: The Einstein equation of state," *Phys.Rev.Lett.* **75** (1995) 1260–1263, arXiv:gr-qc/9504004 [gr-qc].

[46] J. Bekenstein, "Black holes and the second law," *Lett.Nuovo Cim.* **4** (1972) 737–740. http://dx.doi.org/10.1007/BF02757029.

[47] R. Bousso, H. Casini, Z. Fisher, and J. Maldacena, "Proof of a Quantum Bousso Bound," *Phys.Rev.* **D90** (2014) 044002, arXiv:1404.5635 [hep-th].

[48] R. Bousso, H. Casini, Z. Fisher, and J. Maldacena, "Entropy on a null surface for interacting quantum field theories and the Bousso bound," arXiv:1406.4545 [hep-th].

[49] R. Bousso, "The Holographic principle," *Rev.Mod.Phys.* **74** (2002) 825–874, arXiv:hep-th/0203101 [hep-th].

[50] M. Bañados, C. Teitelboim, and J. Zanelli, "Black hole in three-dimensional space-time," *Phys.Rev.Lett.* **69** (1992) 1849–1851, arXiv:hep-th/9204099 [hep-th].

[51] C. Martinez, C. Teitelboim, and J. Zanelli, "Charged rotating black hole in three space-time dimensions," *Phys.Rev.* **D61** (2000) 104013, arXiv:hep-th/9912259 [hep-th].

[52] M. Srednicki, "Entropy and area," *Phys.Rev.Lett.* **71** (1993) 666–669, arXiv:hep-th/9303048 [hep-th].

[53] G. Evenbly and G. Vidal, "Tensor Network States and Geometry," *J.Stat.Phys.* **145** (2011) 891–918, arXiv:1106.1082 [quant-ph].

[54] A. Almheiri, X. Dong, and D. Harlow, "Bulk Locality and Quantum Error Correction in AdS/CFT," arXiv:1411.7041 [hep-th].

[55] G. Evenbly and G. Vidal, "Tensor network renormalization yields the multi-scale entanglement renormalization ansatz," 1502.05385 [cond-mat.str-el].

[56] J. Molina-Vilaplana and J. Prior, "Entanglement, Tensor Networks and Black Hole Horizons," *Gen.Rel.Grav.* **46** (2014) 1823, arXiv:1403.5395 [hep-th].

[57] J. M. Bardeen, B. Carter, and S. Hawking, "The four laws of black hole mechanics," *Commun.Math.Phys.* **31** (1973) 161–170.

[58] J. D. Bekenstein, "Black holes and entropy," *Phys.Rev.* **D7** (1973) 2333–2346.

[59] L. Susskind, "The world as a hologram," *J.Math.Phys.* **36** (1995) 6377–6396, arXiv:hep-th/9409089 [hep-th].

[60] G. 't Hooft, "Dimensional reduction in quantum gravity," arXiv:gr-qc/9310026 [gr-qc].

[61] R. Bousso, "The Holographic principle," *Rev.Mod.Phys.* **74** (2002) 825–874, arXiv:hep-th/0203101 [hep-th].

[62] T. Faulkner, M. Guica, T. Hartman, R. C. Myers, and M. Van Raamsdonk, "Gravitation from Entanglement in Holographic CFTs," *JHEP* **1403** (2014) 051, arXiv:1312.7856 [hep-th].

[63] B. Swingle, "Constructing holographic spacetimes using entanglement renormalization," arXiv:1209.3304 [hep-th].

[64] N. Bao, C. Cao, S. M. Carroll, A. Chatwin-Davies, N. Hunter-Jones, J. Pollack, and G. N. Remmen, "Consistency conditions for an AdS multiscale entanglement renormalization ansatz correspondence," *Phys. Rev.* **D91** (2015) 125036, arXiv:1504.06632 [hep-th].

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

[66] P. Hayden and J. Preskill, "Black holes as mirrors: Quantum information in random subsystems," *JHEP* **0709** (2007) 120, arXiv:0708.4025 [hep-th].

[67] A. Almheiri, D. Marolf, J. Polchinski, and J. Sully, "Black Holes: Complementarity or Firewalls?," *JHEP* **1302** (2013) 062, arXiv:1207.3123 [hep-th].

[68] S. L. Braunstein, "Black hole entropy as entropy of entanglement, or it's curtains for the equivalence principle," arXiv:0907.1190v1 [quant-ph].

[69] K. Jensen and A. Karch, "Holographic Dual of an Einstein-Podolsky-Rosen Pair has a Wormhole," *Phys.Rev.Lett.* **111** no. 21, (2013) 211602, arXiv:1307.1132 [hep-th].

[70] J. Sonner, "Holographic Schwinger Effect and the Geometry of Entanglement," *Phys.Rev.Lett.* **111** no. 21, (2013) 211603, arXiv:1307.6850 [hep-th].

[71] A. Einstein, B. Podolsky, and N. Rosen, "Can quantum mechanical description of physical reality be considered complete?," *Phys.Rev.* **47** (1935) 777–780.

[72] J. Bell, "On the Einstein-Podolsky-Rosen paradox," *Physics* **1** (1964) 195–200.

[73] A. Einstein and N. Rosen, "The Particle Problem in the General Theory of Relativity," *Phys.Rev.* **48** (1935) 73–77.

[74] M. Kruskal, "Maximal extension of Schwarzschild metric," *Phys.Rev.* **119** (1960) 1743–1745.

[75] R. W. Fuller and J. A. Wheeler, "Causality and Multiply Connected Space-Time," *Phys.Rev.* **128** (1962) 919–929.

[76] W. Wootters and W. Zurek, "A single quantum cannot be cloned," *Nature* **299** (1982) 802–803.

[77] R. Geroch, *Singularities in the Spacetime of General Relativity: Their Definition, Existence, and Local Characterization.* PhD thesis, Princeton University, Princeton, 1967.

[78] R. Geroch and G. T. Horowitz, "Global structure of spacetimes," in *General Relativity: An Einstein Centenary Survey*, S. W. Hawking and W. Israel, eds., pp. 212–293. Cambridge University Press, 1979.

[79] F. Tipler, "Singularities and Causality Violation," *Annals Phys.* **108** (1977) 1–36.

[80] F. Tipler, "Singularities and causality violation," *General Relativity and Gravitation* **10** no. 12, (1979) 983–984.

[81] P. Håjíček, "Causality in Non-Hausdorff Space-Times," *Comm. Math. Phys.* **21** no. 1, (1971) 75–84.

[82] J. L. Friedman, K. Schleich, and D. M. Witt, "Topological censorship," *Phys.Rev.Lett.* **71** (1993) 1486–1489, arXiv:gr-qc/9305017 [gr-qc].

[83] G. Galloway, K. Schleich, D. Witt, and E. Woolgar, "Topological censorship and higher genus black holes," *Phys.Rev.* **D60** (1999) 104039, arXiv:gr-qc/9902061 [gr-qc].

[84] D. Dieks, "Communication by EPR devices," *Phys.Lett.* **A92** (1982) 271–272.

[85] J. M. Maldacena, "Eternal black holes in anti-de Sitter," *JHEP* **0304** (2003) 021, arXiv:hep-th/0106112 [hep-th].

[86] M. Van Raamsdonk, "Building up spacetime with quantum entanglement," *Gen.Rel.Grav.* **42** (2010) 2323–2329, arXiv:1005.3035 [hep-th].

[87] G. T. Horowitz and D. Marolf, "A new approach to string cosmology," *JHEP* **9807** (1998) 014, arXiv:hep-th/9805207 [hep-th].

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

[89] S. Hawking and G. Ellis, *The Large Scale Structure of Space-Time.* Cambridge University Press, 1973.

[90] R. P. Geroch, "Domain of Dependence," *J.Math.Phys.* **11** (1970) 437–439.

[91] V. Rubakov, "The null energy condition and its violation," *Phys.Usp.* **57** (2014) 128–142, arXiv:1401.4024 [hep-th].

[92] T. Roman, "Quantum Stress Energy Tensors and the Weak Energy Condition," *Phys.Rev.* **D33** (1986) 3526–3533.

[93] M. Morris, K. Thorne, and U. Yurtsever, "Wormholes, Time Machines, and the Weak Energy Condition," *Phys.Rev.Lett.* **61** (1988) 1446–1449.

[94] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, "Quantum entanglement," *Rev.Mod.Phys.* **81** (2009) 865–942, arXiv:quant-ph/0702225 [quant-ph].

[95] K. S. Thorne, "Closed timelike curves," in *General Relativity and Gravitation 1992*, R. Gleiser, C. Kozameh, and O. Moreschi, eds., pp. 295–316. Taylor & Francis, 1993.

[96] C. Cheung and G. N. Remmen, "Infrared Consistency and the Weak Gravity Conjecture," *JHEP* **1412** (2014) 087, arXiv:1407.7865 [hep-th].

[97] V. Frolov and A. Zelnikov, *Introduction to Black Hole Physics.* Oxford University Press, 2011.

[98] A. Adams, N. Arkani-Hamed, S. Dubovsky, A. Nicolis, and R. Rattazzi, "Causality, analyticity and an IR obstruction to UV completion," *JHEP* **0610** (2006) 014, arXiv:hep-th/0602178 [hep-th].

[99] S. Dubovsky and S. Sibiryakov, "Spontaneous breaking of Lorentz invariance, black holes and perpetuum mobile of the 2nd kind," *Phys.Lett.* **B638** (2006) 509–514, arXiv:hep-th/0603158 [hep-th].

[100] T. A. Brun, M. M. Wilde, and A. Winter, "Quantum state cloning using deutschian closed timelike curves," *Phys. Rev. Lett.* **111** (Nov, 2013) 190401, arXiv:1306.1795 [quant-ph].

[101] G. 't Hooft, "On the Quantum Structure of a Black Hole," *Nucl. Phys.* **B256** (1985) 727.

[102] G. 't Hooft, "The black hole interpretation of string theory," *Nucl. Phys.* **B335** (1990) 138–154.

[103] K. Papadodimas and S. Raju, "An Infalling Observer in AdS/CFT," *JHEP* **10** (2013) 212, arXiv:1211.6767 [hep-th].

[104] K. Papadodimas and S. Raju, "Black Hole Interior in the Holographic Correspondence and the Information Paradox," *Phys. Rev. Lett.* **112** no. 5, (2014) 051301, arXiv:1310.6334 [hep-th].

[105] K. Papadodimas and S. Raju, "State-Dependent Bulk-Boundary Maps and Black Hole Complementarity," *Phys. Rev.* **D89** no. 8, (2014) 086010, arXiv:1310.6335 [hep-th].

[106] K. Papadodimas and S. Raju, "Local Operators in the Eternal Black Hole," arXiv:1502.06692 [hep-th].

[107] K. Papadodimas and S. Raju, "Comments on the Necessity and Implications of State-Dependence in the Black Hole Interior," arXiv:1503.08825 [hep-th].

[108] R. Bousso, "Firewalls from double purity," *Phys. Rev.* **D88** no. 8, (2013) 084035, arXiv:1308.2665 [hep-th].

[109] D. Marolf and J. Polchinski, "Violations of the Born rule in cool state-dependent horizons," arXiv:1506.01337 [hep-th].

[110] J. C. Baez and J. Vicary, "Wormholes and Entanglement," *Class. Quant. Grav.* **31** no. 21, (2014) 214007, arXiv:1401.3416 [gr-qc].

[111] H. Gharibyan and R. F. Penna, "Are entangled particles connected by wormholes? Evidence for the ER=EPR conjecture from entropy inequalities," *Phys. Rev.* **D89** no. 6, (2014) 066001, arXiv:1308.0289 [hep-th].

[112] G. Mandal, R. Sinha, and N. Sorokhaibam, "The inside outs of AdS$_3$/CFT$_2$: exact AdS wormholes with entangled CFT duals," *JHEP* **01** (2015) 036, arXiv:1405.6695 [hep-th].

[113] N. Bao, J. Pollack, and G. N. Remmen, "Splitting Spacetime and Cloning Qubits: Linking No-Go Theorems across the ER=EPR Duality," arXiv:1506.08203 [hep-th].

[114] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information.* Cambridge University Press, 2010.

[115] J. Griffiths and J. Podolský, *Exact Space-Times in Einstein's General Relativity.* Cambridge University Press, 2009.

[116] K. Lake and R. C. Roeder, "Effects of a Nonvanishing Cosmological Constant on the Spherically Symmetric Vacuum Manifold," *Phys. Rev.* **D15** (1977) 3513–3519.

[117] S. Hemming and E. Keski-Vakkuri, "Hawking radiation from AdS black holes," *Phys. Rev.* **D64** (2001) 044006, arXiv:gr-qc/0005115 [gr-qc].

[118] G. T. Horowitz and V. E. Hubeny, "Quasinormal modes of AdS black holes and the approach to thermal equilibrium," *Phys. Rev.* **D62** (2000) 024027, arXiv:hep-th/9909056 [hep-th].

[119] L. Fidkowski, V. Hubeny, M. Kleban, and S. Shenker, "The black hole singularity in AdS/CFT," *JHEP* **02** (2004) 014, arXiv:hep-th/0306170 [hep-th].

[120] J. L. Friedman, K. Schleich, and D. M. Witt, "Topological censorship," *Phys. Rev. Lett.* **71** (1993) 1486–1489, arXiv:gr-qc/9305017 [gr-qc].

[121] V. Coffman, J. Kundu, and W. K. Wootters, "Distributed entanglement," *Phys. Rev.* **A61** (2000) 052306, arXiv:quant-ph/9907047 [quant-ph].

[122] E. H. Lieb and M. B. Ruskai, "Proof of the strong subadditivity of quantum-mechanical entropy," *J. Math. Phys.* **14** (1973) 1938–1941.

[123] C. Holzhey, F. Larsen, and F. Wilczek, "Geometric and renormalized entropy in conformal field theory," *Nucl. Phys.* **B424** (1994) 443–467, arXiv:hep-th/9403108 [hep-th].

[124] P. Calabrese and J. L. Cardy, "Entanglement entropy and quantum field theory," *J. Stat. Mech.* **0406** (2004) P06002, arXiv:hep-th/0405152 [hep-th].

[125] R. Bousso, H. Casini, Z. Fisher, and J. Maldacena, "Entropy on a null surface for interacting quantum field theories and the Bousso bound," *Phys. Rev.* **D91** (2015) 084030, arXiv:1406.4545 [hep-th].

[126] P. Hayden, M. Headrick, and A. Maloney, "Holographic Mutual Information is Monogamous," *Phys. Rev.* **D87** (2013) 046003, arXiv:1107.2940 [hep-th].

[127] N. Bao, S. Nezami, H. Ooguri, B. Stoica, J. Sully, and M. Walter, "The Holographic Entropy Cone," *JHEP* **09** (2015) 130, arXiv:1505.07839 [hep-th].

[128] L. Susskind, "The world as a hologram," *J.Math.Phys.* **36** (1995) 6377–6396, arXiv:hep-th/9409089 [hep-th].

[129] L. Susskind, "Copenhagen vs Everett, Teleportation, and ER=EPR," arXiv:1604.02589 [hep-th].

[130] N. Bao, J. Pollack, and G. N. Remmen, "Splitting Spacetime and Cloning Qubits: Linking No-Go Theorems across the ER=EPR Duality," *Fortsch. Phys.* **63** (2015) 705–710, arXiv:1506.08203 [hep-th].

[131] N. Bao, J. Pollack, and G. N. Remmen, "Wormhole and Entanglement (Non-)Detection in the ER=EPR Correspondence," *JHEP* **11** (2015) 126, arXiv:1509.05426 [hep-th].

[132] S. W. Hawking, "Black holes in general relativity," *Commun. Math. Phys.* **25** (1972) 152–166.

[133] S. Hawking and G. Ellis, *The Large Scale Structure of Space-Time.* Cambridge University Press, 1973.

[134] A. Królak, "Definitions of black holes without use of the boundary at infinity," *General Relativity and Gravitation* **14** 793–801.

[135] R. Bousso, H. Casini, Z. Fisher, and J. Maldacena, "Proof of a Quantum Bousso Bound," *Phys. Rev.* **D90** (2014) 044002, arXiv:1404.5635 [hep-th].

[136] R. Bousso and N. Engelhardt, "New Area Law in General Relativity," *Phys. Rev. Lett.* **115** (2015) 081301, arXiv:1504.07627 [hep-th].

[137] V. E. Hubeny, M. Rangamani, and T. Takayanagi, "A covariant holographic entanglement entropy proposal," *JHEP* **07** (2007) 062, arXiv:0705.0016 [hep-th].

[138] A. C. Wall, "Maximin Surfaces, and the Strong Subadditivity of the Covariant Holographic Entanglement Entropy," *Class. Quant. Grav.* **31** (2014) 225007, arXiv:1211.3494 [hep-th].

[139] V. Balasubramanian, P. Hayden, A. Maloney, D. Marolf, and S. F. Ross, "Multiboundary Wormholes and Holographic Entanglement," *Class. Quant. Grav.* **31** (2014) 185015, arXiv:1406.2663 [hep-th].

[140] R. P. Geroch, "Domain of dependence," *J. Math. Phys.* **11** (1970) 437.

[141] M. Headrick, V. E. Hubeny, A. Lawrence, and M. Rangamani, "Causality & holographic entanglement entropy," *JHEP* **12** (2014) 162, arXiv:1408.6300 [hep-th].

[142] S. M. Carroll, *Spacetime and Geometry: An Introduction to General Relativity.* Addison Wesley, 2004.

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

[144] Y. Nomura, "Physical Theories, Eternal Inflation, and Quantum Universe," *JHEP* **11** (2011) 063, arXiv:1104.2324 [hep-th].

[145] T. Bunch and P. Davies, "Quantum Field Theory in de Sitter Space: Renormalization by Point Splitting," *Proc. Roy. Soc. Lond.* **A360** (1978) 117.

[146] T. Bunch and P. Davies, "Nonconformal Renormalized Stress Tensors in Robertson-Walker Space-Times," *J. Phys.* **A11** (1978) 1315.

[147] J. Hartle and S. Hawking, "Path Integral Derivation of Black Hole Radiance,"
Phys. Rev. **D13** (1976) 2188.

[148] G. Gibbons and S. Hawking, "Cosmological event horizons, thermodynamics, and
particle creation," *Phys. Rev.* **D15** (1977) 2738.

[149] T. Banks, "Cosmological breaking of supersymmetry? or Little lambda goes back
to the future 2," *Int. J. Mod. Phys.* **A16** (2001) 910, arXiv:hep-th/0007146
[hep-th].

[150] T. Banks and W. Fischler, "M theory observables for cosmological space-times,"
arXiv:hep-th/0102077 [hep-th].

[151] D. H. Lyth and A. R. Liddle, *The Primordial Density Perturbation: Cosmology,
Inflation and the Origin of Structure.* Cambridge University Press, Revised edition
2009.

[152] S. Dodelson, *Modern Cosmology.* Academic Press, San Diego, CA, 2003.

[153] D. Baumann, "TASI Lectures on Inflation," arXiv:0907.5424 [hep-th].

[154] A. Vilenkin, "The Birth of Inflationary Universes," *Phys. Rev.* **D27** (1983) 2848.

[155] A. S. Goncharov and A. D. Linde, "Global structure of the inflationary universe,"
Zh. Eksp. Teor. Fiz. **92** (1987) 1137.

[156] A. Goncharov, A. D. Linde, and V. F. Mukhanov, "The Global Structure of the
Inflationary Universe," *Int. J. Mod. Phys.* **A2** (1987) 561.

[157] S. Weinberg, "Anthropic Bound on the Cosmological Constant," *Phys. Rev. Lett.*
59 (1987) 2607.

[158] R. Bousso and J. Polchinski, "Quantization of four form fluxes and dynamical
neutralization of the cosmological constant," *JHEP* **06** (2000) 006,
arXiv:hep-th/0004134 [hep-th].

[159] S. Kachru, R. Kallosh, A. D. Linde, and S. P. Trivedi, "De Sitter vacua in string
theory," *Phys. Rev.* **D68** (2003) 046005, arXiv:hep-th/0301240 [hep-th].

[160] L. Susskind, "The Anthropic landscape of string theory," in *The Davis Meeting On
Cosmic Inflation.* Mar., 2003. arXiv:hep-th/0302219 [hep-th].

[161] F. Denef and M. R. Douglas, "Distributions of flux vacua," *JHEP* **05** (2004) 072,
arXiv:hep-th/0404116 [hep-th].

[162] K.-M. Lee and E. J. Weinberg, "Decay of the True Vacuum in Curved Space-time,"
Phys. Rev. **D36** (1987) 1088.

[163] A. Aguirre, S. M. Carroll, and M. C. Johnson, "Out of equilibrium: understanding cosmological evolution to lower-entropy states," *JCAP* **02** (2012) 024, arXiv:1108.0417 [hep-th].

[164] L. Dyson, M. Kleban, and L. Susskind, "Disturbing implications of a cosmological constant," *JHEP* **10** (2002) 011.

[165] A. Albrecht and L. Sorbo, "Can the universe afford inflation?," *Phys. Rev.* **D70** (2004) 063528, arXiv:hep-th/0405270 [hep-th].

[166] R. Bousso and B. Freivogel, "A Paradox in the global description of the multiverse," *JHEP* **06** (2007) 018, arXiv:hep-th/0610132 [hep-th].

[167] T. Banks, W. Fischler, and S. Paban, "Recurrent nightmares? Measurement theory in de Sitter space," *JHEP* **12** (2002) 062, arXiv:hep-th/0210160 [hep-th].

[168] M. Spradlin, A. Strominger, and A. Volovich, "Les Houches lectures on de Sitter space," arXiv:hep-th/0110007 [hep-th].

[169] C. R. Stephens, G. 't Hooft, and B. F. Whiting, "Black hole evaporation without information loss," *Class. Quant. Grav.* **11** (1994) 621, arXiv:gr-qc/9310006 [gr-qc].

[170] M. K. Parikh, I. Savonije, and E. P. Verlinde, "Elliptic de Sitter space: dS/Z(2)," *Phys. Rev.* **D67** (2003) 064005, arXiv:hep-th/0209120 [hep-th].

[171] H. Everett, "Relative state formulation of quantum mechanics," *Rev. Mod. Phys.* **29** (1957) 454.

[172] D. Wallace, *The Emergent Multiverse.* Oxford University Press, Oxford, 2012.

[173] A. Elby and J. Bub, "Triorthogonal uniqueness theorem and its relevance to the interpretation of quantum mechanics," *Phys. Rev.* **A49** (1994) 4213.

[174] W. H. Zurek, "Preferred states, predictability, classicality and the environment-induced decoherence," *Prog. Theor. Phys.* **89** (1993) 281.

[175] W. H. Zurek, "Decoherence, Einselection, and the existential interpretation: The Rough guide," *Phil. Trans. Roy. Soc. Lond.* **A356** (1998) 1793, arXiv:quant-ph/9805065 [quant-ph].

[176] W. H. Zurek, "Decoherence, einselection, and the quantum origins of the classical," *Rev. Mod. Phys.* **75** (2003) 715, arXiv:quant-ph/0105127.

[177] S. Khlebnikov and M. Kruczenski, "Thermalization of isolated quantum systems," arXiv:1312.4612 [cond-mat.stat-mech].

[178] D. N. Page and W. K. Wootters, "Evolution without evolution: Dynamics described by stationary observables," *Phys. Rev.* **D27** (1983) 2885.

[179] D. N. Page, "The Lifetime of the universe," *J. Korean Phys. Soc.* **49** (2006) 711, arXiv:hep-th/0510003 [hep-th].

[180] M. Davenport and K. D. Olum, "Are there Boltzmann brains in the vacuum?," arXiv:1008.0808 [hep-th].

[181] G. Lindblad, "On the Generators of Quantum Dynamical Semigroups," *Commun. Math. Phys.* **48** (1976) 119.

[182] F. Marquardt and A. Püttmann, "Introduction to dissipation and decoherence in quantum systems," arXiv:0809.4403 [quant-ph].

[183] W. Zurek, "Environment induced superselection rules," *Phys. Rev.* **D26** (1982) 1862.

[184] N. D. Birrell and P. C. W. Davies, *Quantum Fields in Curved Space.* Cambridge Monographs on Mathematical Physics. Cambridge University Press, 1984.

[185] B. Allen, "Vacuum States in de Sitter Space," *Phys. Rev.* **D32** (1985) 3136.

[186] D. N. Page and X. Wu, "Massless Scalar Field Vacuum in de Sitter Spacetime," *JCAP* **11** (2012) 51, arXiv:1204.4462 [hep-th].

[187] P. Candelas and D. Raine, "General Relativistic Quantum Field Theory-An Exactly Soluble Model," *Phys. Rev.* **D12** (1975) 965.

[188] J. Géhéniau and C. Schomblond, "Fonctions de green dans l'univers de de sitter," *Acad. R. Belg. Bull. Cl. Sci.* **54** (1968) 1147.

[189] C. Schomblond and P. Spindel, "Unicity Conditions of the Scalar Field Propagator Delta(1) (x,y) in de Sitter Universe," *Annales Poincare Phys.Theor.* **25** (1976) 67.

[190] N. Chernikov and E. Tagirov, "Quantum theory of scalar fields in de Sitter space-time," *Annales Poincare Phys.Theor.* **A9** (1968) 109.

[191] E. Tagirov, "Consequences of field quantization in de Sitter type cosmological models," *Annals Phys.* **76** (1973) 561.

[192] E. Mottola, "Particle Creation in de Sitter Space," *Phys. Rev.* **D31** (1985) 754.

[193] R. Bousso, A. Maloney, and A. Strominger, "Conformal vacua and entropy in de sitter space," *Phys. Rev.* **D65** (2002) 104039.

[194] P. R. Anderson, W. Eaker, S. Habib, C. Molina-Paris, and E. Mottola, "Attractor states and infrared scaling in de Sitter space," *Phys. Rev.* **D62** (2000) 124019, arXiv:gr-qc/0005102 [gr-qc].

[195] S. Hollands, "Massless interacting quantum fields in deSitter spacetime," *Annales Henri Poincare* **13** (2012) 1039, arXiv:1105.1996 [gr-qc].

[196] B. Garbrecht and G. Rigopoulos, "Self Regulation of Infrared Correlations for Massless Scalar Fields during Inflation," *Phys. Rev.* **D84** (2011) 063516, arXiv:1105.0418 [hep-th].

[197] B. Garbrecht, G. Rigopoulos, and Y. Zhu, "Infrared Correlations in de Sitter Space: Field Theoretic vs. Stochastic Approach," *Phys. Rev.* **D89** (2014) 063506, arXiv:1310.0367 [hep-th].

[198] Y. Nomura, "Quantum Mechanics, Spacetime Locality, and Gravity," *Found. Phys.* **43** (2013) 978, arXiv:1110.4630 [hep-th].

[199] J. D. Bekenstein, "Black holes and entropy," *Phys. Rev.* **D7** (1973) 2333.

[200] S. Hawking, "Gravitational radiation from colliding black holes," *Phys. Rev. Lett.* **26** (1971) 1344.

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

[202] T. Banks, "Some thoughts on the quantum theory of stable de Sitter space," arXiv:hep-th/0503066 [hep-th].

[203] S. B. Giddings and D. Marolf, "A Global picture of quantum de Sitter space," *Phys. Rev.* **D76** (2007) 064023, arXiv:0705.1178 [hep-th].

[204] S. R. Coleman, "The Fate of the False Vacuum. 1. Semiclassical Theory," *Phys. Rev.* **D15** (1977) 2929.

[205] S. R. Coleman and F. De Luccia, "Gravitational Effects on and of Vacuum Decay," *Phys. Rev.* **D21** (1980) 3305.

[206] L. Susskind, "The Census taker's hat," arXiv:0710.1129 [hep-th].

[207] Y. Sekino and L. Susskind, "Census Taking in the Hat: FRW/CFT Duality," *Phys. Rev.* **D80** (2009) 083531, arXiv:0908.3844 [hep-th].

[208] R. Bousso and L. Susskind, "The Multiverse Interpretation of Quantum Mechanics," *Phys. Rev.* **D85** (2012) 045007, arXiv:1105.3796 [hep-th].

[209] D. N. Page, "Is our universe likely to decay within 20 billion years?," *Phys. Rev.* **D78** (2008) 063535, arXiv:hep-th/0610079 [hep-th].

[210] D. N. Page, "Susskind's challenge to the Hartle-Hawking no-boundary proposal and possible resolutions," *JCAP* **1** (2007) 004, arXiv:hep-th/0610199 [hep-th].

[211] D. N. Page, "Return of the Boltzmann Brains," *Phys. Rev.* **D78** (2008) 063536, arXiv:hep-th/0611158 [hep-th].

[212] D. N. Page, "Is our universe decaying at an astronomical rate?," *Phys. Lett.* **B669** (2008) 197, arXiv:hep-th/0612137 [hep-th].

[213] D. N. Page, "Possible Anthropic Support for a Decaying Universe: A Cosmic Doomsday Argument," arXiv:0907.4153 [hep-th].

[214] J. R. Gott III, "Boltzmann Brains: I'd Rather See Than Be One," arXiv:0802.0233 [gr-qc].

[215] S. Aaronson, "The Ghost in the Quantum Turing Machine," arXiv:1306.0159 [quant-ph].

[216] S. M. Carroll, "What if Time Really Exists?," arXiv:0811.3772 [gr-qc].

[217] K. K. Boddy and S. M. Carroll, "Can the Higgs Boson Save Us From the Menace of the Boltzmann Brains?," arXiv:1308.4686 [hep-ph].

[218] J. Garriga and A. Vilenkin, "Recycling universe," *Phys. Rev.* **D57** (1998) 2230, arXiv:astro-ph/9707292 [astro-ph].

[219] A. D. Linde, "Sinks in the Landscape, Boltzmann Brains, and the Cosmological Constant Problem," *JCAP* **1** (2007) 022, arXiv:hep-th/0611043 [hep-th].

[220] D. Polarski and A. A. Starobinsky, "Semiclassicality and decoherence of cosmological perturbations," *Class. Quant. Grav.* **13** (1996) 377, arXiv:gr-qc/9504030 [gr-qc].

[221] F. C. Lombardo and D. Lopez Nacir, "Decoherence during inflation: The Generation of classical inhomogeneities," *Phys. Rev.* **D72** (2005) 063506, arXiv:gr-qc/0506051 [gr-qc].

[222] P. Martineau, "On the decoherence of primordial fluctuations during inflation," *Class. Quant. Grav.* **24** (2007) 5817, arXiv:astro-ph/0601134 [astro-ph].

[223] C. P. Burgess, R. Holman, and D. Hoover, "Decoherence of inflationary primordial fluctuations," *Phys. Rev.* **D77** (2008) 063534, arXiv:astro-ph/0601646 [astro-ph].

[224] C. Kiefer, I. Lohmar, D. Polarski, and A. A. Starobinsky, "Pointer states for primordial fluctuations in inflationary cosmology," *Class. Quant. Grav.* **24** (2007) 1699, arXiv:astro-ph/0610700 [astro-ph].

[225] T. Prokopec and G. I. Rigopoulos, "Decoherence from Isocurvature perturbations in Inflation," *JCAP* **11** (2007) 029, arXiv:astro-ph/0612067 [astro-ph].

[226] P. Creminelli, S. Dubovsky, A. Nicolis, L. Senatore, and M. Zaldarriaga, "The Phase Transition to Slow-roll Eternal Inflation," *JHEP* **09** (2008) 036, arXiv:0802.1067 [hep-th].

[227] S. Dubovsky, L. Senatore, and G. Villadoro, "Universality of the Volume Bound in Slow-Roll Eternal Inflation," *JHEP* **05** (2012) 035, arXiv:1111.1725 [hep-th].

[228] E. J. Martinec and W. E. Moore, "Modeling Quantum Gravity Effects in Inflation," arXiv:1401.7681 [hep-th].

[229] **BICEP2** Collaboration, P. Ade *et al.*, "BICEP2 I: Detection Of B-mode Polarization at Degree Angular Scales," *Phys. Rev. Lett.* **112** (2014) 241101, arXiv:1403.3985 [astro-ph.CO].

[230] R. Bousso, "Proliferation of de Sitter space," *Phys. Rev.* **D58** (1998) 083511, arXiv:hep-th/9805081 [hep-th].

[231] D. Bohm, "A Suggested Interpretation of the Quantum Theory in Terms of 'Hidden' Variables. I," *Phys. Rev.* **85** (1952) 166.

[232] D. Bohm, "A Suggested Interpretation of the Quantum Theory in Terms of 'Hidden' Variables. II," *Phys. Rev.* **85** (1952) 180.

[233] D. Dürr, S. Goldstein, R. Tumulka, and N. Zanghì, "Bohmian mechanics and quantum field theory," *Phys. Rev. Lett.* **93** (2004) 090402.

[234] W. Struyve, "Pilot-wave theory and quantum fields," *Rept. Prog. Phys.* **73** (2010) 106001, arXiv:0707.3685 [quant-ph].

[235] S. Goldstein, W. Struyve, and R. Tumulka, "The Bohmian Approach to the Problems of Cosmological Quantum Fluctuations," arXiv:1508.01017 [gr-qc].

[236] G. Ghirardi, A. Rimini, and T. Weber, "A model for a unified quantum description of macroscopic and microscopic systems," in *Quantum Probability and Applications II*, pp. 223–232. Springer, 1985.

[237] G. C. Ghirardi, A. Rimini, and T. Weber, "Unified dynamics for microscopic and macroscopic systems," *Phys. Rev.* **D34** (1986) 470.

[238] S. M. Carroll, "In What Sense Is the Early Universe Fine-Tuned?," 2014. arXiv:1406.3057 [astro-ph.CO]. https://inspirehep.net/record/1300332/files/arXiv:1406.3057.pdf.

[239] A. H. Guth, "Inflationary universe: A possible solution to the horizon and flatness problems," *Phys. Rev.* **D23** (Jan., 1981) 347–356.

[240] A. D. Linde, "A new inflationary universe scenario: A possible solution of the horizon, flatness, homogeneity, isotropy and primordial monopole problems," *Phys. Lett.* **B108** (Feb., 1982) 389–393.

[241] A. Albrecht and P. J. Steinhardt, "Cosmology for grand unified theories with

radiatively induced symmetry breaking," *Phys. Rev. Lett.* **48** (Apr., 1982) 1220–1223.

[242] C. Burgess, R. Holman, G. Tasinato, and M. Williams, "EFT Beyond the Horizon: Stochastic Inflation and How Primordial Quantum Fluctuations Go Classical," arXiv:1408.5002 [hep-th].

[243] J. Liu, C.-M. Sou, and Y. Wang, "Cosmic Decoherence: Massive Fields," *JHEP* **10** (2016) 072, arXiv:1608.07909 [hep-th].

[244] A. D. Linde, "Scalar field fluctuations in the expanding universe and the new inflationary universe scenario," *Phys. Lett.* **B116** (Oct., 1982) 335–339.

[245] A. A. Starobinsky, "Dynamics of phase transition in the new inflationary universe scenario and generation of perturbations," *Phys. Lett.* **B117** (Nov., 1982) 175–178.

[246] A. D. Linde, "Eternally Existing Selfreproducing Chaotic Inflationary Universe," *Phys. Lett.* **B175** (1986) 395–400.

[247] K. K. Boddy, S. M. Carroll, and J. Pollack, "De Sitter Space Without Dynamical Quantum Fluctuations," *Found. Phys.* (Mar., 2016) 1–34, arXiv:1405.0298 [hep-th].

[248] E. Nelson, "Quantum Decoherence During Inflation from Gravitational Nonlinearities," *JCAP* **1603** (2016) 022, arXiv:1601.03734 [gr-qc].

[249] A. Vilenkin and L. H. Ford, "Gravitational effects upon cosmological phase transitions," *Phys. Rev.* **D26** (1982) 1231.

[250] L. P. Grishchuk and Yu. V. Sidorov, "Squeezed quantum states of relic gravitons and primordial density fluctuations," *Phys. Rev.* **D42** (1990) 3413–3421.

[251] A. Albrecht, P. Ferreira, M. Joyce, and T. Prokopec, "Inflation and squeezed quantum states," *Phys. Rev.* **D50** (1994) 4807–4820, arXiv:astro-ph/9303001 [astro-ph].

[252] J. M. Maldacena, "Non-Gaussian features of primordial fluctuations in single field inflationary models," *JHEP* **05** (2003) 013, arXiv:astro-ph/0210603 [astro-ph].

[253] **Planck** Collaboration, P. A. R. Ade *et al.*, "Planck 2015 results. XIII. Cosmological parameters," *Astron. Astrophys.* **594** (2016) A13, arXiv:1502.01589 [astro-ph.CO].

[254] E. Pajer, F. Schmidt, and M. Zaldarriaga, "The Observed Squeezed Limit of Cosmological Three-Point Functions," *Phys. Rev.* **D88** no. 8, (2013) 083502, arXiv:1305.0824 [astro-ph.CO].

[255] T. Tanaka and Y. Urakawa, "Dominance of gauge artifact in the consistency relation for the primordial bispectrum," *JCAP* **1105** (2011) 014, arXiv:1103.1251 [astro-ph.CO].

[256] C. T. Sebens and S. M. Carroll, "Self-Locating Uncertainty and the Origin of Probability in Everettian Quantum Mechanics," arXiv:1405.7577 [quant-ph].

[257] S. Winitzki, "Predictions in eternal inflation," *Lect. Notes Phys.* **738** (2008) 157–191, arXiv:gr-qc/0612164 [gr-qc].

[258] B. Freivogel, "Making predictions in the multiverse," *Class. Quant. Grav.* **28** (2011) 204007, arXiv:1105.0244 [hep-th].

[259] Y. Nomura, "The Static Quantum Multiverse," *Phys. Rev.* **D86** (2012) 083505, arXiv:1205.5550 [hep-th].

[260] S. W. Hawking, Phys. Rev. D **72**, 084013 (2005) doi:10.1103/PhysRevD.72.084013 [hep-th/0507171].

[261] A. Almheiri, D. Marolf, J. Polchinski and J. Sully, JHEP **1302**, 062 (2013) doi:10.1007/JHEP02(2013)062 [arXiv:1207.3123 [hep-th]].

[262] A. Almheiri, D. Marolf, J. Polchinski, D. Stanford and J. Sully, JHEP **1309**, 018 (2013) doi:10.1007/JHEP09(2013)018 [arXiv:1304.6483 [hep-th]].

[263] L. Susskind, L. Thorlacius and J. Uglum, Phys. Rev. D **48**, 3743 (1993) doi:10.1103/PhysRevD.48.3743 [hep-th/9306069].

[264] P. Chen, Y. C. Ong and D. h. Yeom, Phys. Rept. **603**, 1 (2015) doi:10.1016/j.physrep.2015.10.007 [arXiv:1412.8366 [gr-qc]].

[265] J. Maldacena and L. Susskind, Fortsch. Phys. **61**, 781 (2013) doi:10.1002/prop.201300020 [arXiv:1306.0533 [hep-th]].

[266] S. Lloyd and J. Preskill, JHEP **1408**, 126 (2014) doi:10.1007/JHEP08(2014)126 [arXiv:1308.4209 [hep-th]].

[267] K. Papadodimas and S. Raju, JHEP **1310**, 212 (2013) doi:10.1007/JHEP10(2013)212 [arXiv:1211.6767 [hep-th]].

[268] S. D. Mathur, Fortsch. Phys. **53**, 793 (2005) doi:10.1002/prop.200410203 [hep-th/0502050].

[269] S. B. Giddings, Phys. Rev. D **88**, 064023 (2013) doi:10.1103/PhysRevD.88.064023 [arXiv:1211.7070 [hep-th]].

[270] M. Hotta and A. Sugita, PTEP **2015**, 123B04 (2015) doi:10.1093/ptep/ptv170 [arXiv:1505.05870 [gr-qc]].

[271] Y. Nomura, F. Sanches and S. J. Weinberg, Phys. Rev. Lett. **114**, 201301 (2015) doi:10.1103/PhysRevLett.114.201301 [arXiv:1412.7539 [hep-th]].

[272] Y. Nomura, F. Sanches and S. J. Weinberg, JHEP **1504**, 158 (2015) doi:10.1007/JHEP04(2015)158 [arXiv:1412.7538 [hep-th]].

[273] Y. Nomura and N. Salzetta, arXiv:1602.07673 [hep-th].

[274] D. Harlow and P. Hayden, JHEP **1306**, 085 (2013) doi:10.1007/JHEP06(2013)085 [arXiv:1301.4504 [hep-th]].

[275] C. H. Bennett, H. J. Bernstein, S. Popescu, and B. Schumacher, Physical Review A **53**, 2046 (1996) doi:10.1103/PhysRevA.53.2046 [quant-ph/9511030].

[276] J. Oppenheim and W. G. Unruh, JHEP **1403**, 120 (2014) doi:10.1007/JHEP03(2014)120 [arXiv:1401.1523 [hep-th]].

[277] S. Aaronson, "Computational complexity underpinnings of the Harlow-Hayden argument," unpublished.

[278] K. S. Thorne, R. H. Price and D. A. Macdonald, NEW HAVEN, USA: YALE UNIV. PR. (1986) 367p

[279] Y. Nomura, JHEP **1111**, 063 (2011) doi:10.1007/JHEP11(2011)063 [arXiv:1104.2324 [hep-th]].

[280] P. Hayden and J. Preskill, JHEP **0709**, 120 (2007) doi:10.1088/1126-6708/2007/09/120 [arXiv:0708.4025 [hep-th]].

[281] Y. Sekino and L. Susskind, JHEP **0810**, 065 (2008) doi:10.1088/1126-6708/2008/10/065 [arXiv:0808.2096 [hep-th]].

[282] N. Lashkari, D. Stanford, M. Hastings, T. Osborne and P. Hayden, JHEP **1304**, 022 (2013) doi:10.1007/JHEP04(2013)022 [arXiv:1111.6580 [hep-th]].

[283] S. H. Shenker and D. Stanford, JHEP **1403**, 067 (2014) doi:10.1007/JHEP03(2014)067 [arXiv:1306.0622 [hep-th]].

[284] S. H. Shenker and D. Stanford, JHEP **1505**, 132 (2015) doi:10.1007/JHEP05(2015)132 [arXiv:1412.6087 [hep-th]].

[285] J. Maldacena, S. H. Shenker and D. Stanford, arXiv:1503.01409 [hep-th].

[286] D. N. Page, Phys. Rev. D, **13**, 198 (1976) doi:10.1103/PhysRevD.13.198

[287] D. N. Page, Phys. Rev. D, **14**, 3260 (1976) doi:10.1103/PhysRevD.14.3260

[288] S. B. Giddings, Phys. Lett. B **754**, 39 (2016) doi:10.1016/j.physletb.2015.12.076 [arXiv:1511.08221 [hep-th]].

[289] D. N. Page, Phys. Rev. Lett, **71**, 3743 (1993) 10.1103/PhysRevLett.71.3743 [hep-th/9306083].

[290] D. N. Page, JCAP, **2013**, 028 (2013) 10.1088/1475-7516/2013/09/028 [arXiv:1301.4995 [hep-th]].

[291] F. GSL. Brandao, A. Harrow, and M. Horodecki, arXiv:1208.0692 [quant-ph] (2012).

[292] C. Dankert, R. Cleve, J. Emerson, and E. Livine, Phs. Rev. Lett. A, **80**, 1 (2009).

[293] J. Shore and R. Johnson, IEEE Trans. on Inf. Theory. **26** 1 (1980)

[294] B. Freivogel, R. A. Jefferson, L. Kabir, and I.-S. Yang, Phys. Rev. D 91, 044036 (2015).

[295] S. Raju, arXiv:1604.03095 [hep-th] (2016).

[296] S. D. Mathur and D. Turton, Nucl. Phys. B **884**, 566 (2014) doi:10.1016/j.nuclphysb.2014.05.012 [arXiv:1306.5488 [hep-th]].

[297] S. D. Mathur, arXiv:1506.04342 [hep-th].

[298] G. Dotti and R. J. Gleiser, Class. Quant. Grav. **27**, 185007 (2010) doi:10.1088/0264-9381/27/18/185007 [arXiv:1001.0152 [gr-qc]].

[299] N. Engelhardt and G. T. Horowitz, Phys. Rev. D **93**, no. 2, 026005 (2016) doi:10.1103/PhysRevD.93.026005 [arXiv:1509.07509 [hep-th]].