Title
Locality in Gravity

Author
Sanches, Fabio

Publication Date
2018

Peer reviewed|Thesis/dissertation

Locality in Gravity

by

Fabio Sanches

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Physics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Yasunori Nomura, Chair
Professor Ori Ganor
Professor Richard Borcherds

Summer 2018

Locality in Gravity

Abstract

Locality in Gravity

by

Fabio Sanches

Doctor of Philosophy in Physics

University of California, Berkeley

Professor Yasunori Nomura, Chair

The research presented in this dissertation is primarily focused around the study of locality in quantum gravity. The emergence of locality is intimately tied to many important questions in the field, including the emergence of the bulk in the Anti-de Sitter/Conformal Field Theory correspondence, and to holography in more general spacetimes, as well as the contradictions presented in the AMPS argument of the black hole information paradox. In particular, this thesis starts by studying the gauge redundancy related to the observer dependence in quantum gravity as well as its implications in the distribution of gravitational degrees of freedom. This picture is then applied to the black hole information paradox presenting a picture discussing deviations of local effective field theory around the horizon. Following the apparent fundamental role taken by the holographic entanglement proposal in AdS/CFT, I then present a generalization of this proposal to general spacetimes, proving that the appropriate a generalization satisfies the inequalities associated with von Neumann entropy. This picture is then used to study the Hilbert space structure of such theories. In chapter 4, I also present a model that suppressed isocurvature fluctuations present for interesting parameter ranges for axions in high scale inflation scenarios. Finally, the last chapter concludes with the reconstruction of bulk operators in AdS/CFT, where such a reconstruction has no prior knowledge of bulk geometry as a starting point. The result uses intimate connections to quantum error correction and obtains the bulk conformal metric as a byproduct of the construction.

To my family

Contents

List of Figures

Acknowledgments

There are many people who have made meaningful contributions and been invaluable to me throughout the course of my PhD. It would be hard to provide a complete list of everyone who has had an impact in my work.

I am profoundly grateful to my parents, Diane and Ivo, and my sisters, Mayara and Bruna, who have shown me nothing but unfailing support and encouragement. This thesis would not have been possible without them.

I want to thank my advisor, Yasunori Nomura, for the guidance and motivation I've received throughout my PhD. The knowledge of physics I have learned from him and his mentorship were invaluable throughout my PhD.

I also want to thank my extended family, and friends, you have all made this journey fun at every moment. I especially want to thank Matt and Jason.

Thanks to my fellow BCTP members for the numerous insightful and stimulating discussions and collaborations that have helped me understand many topics and have made my time in Berkeley very enjoyable.

My work was generously supported by the Department of Energy NNSA Stewardship Science Graduate Fellowship.

Chapter 1

Introduction

One of the most important open research fields in theoretical physics concerns the quantum mechanical description of gravity. This thesis presents a collection of papers which study, in some form, the quantum mechanical description of spacetime and its implications, as well as a paper discussing the effects of cosmological inflation in phenomenological aspects of high energy particle physics. The study of quantum gravity is not a recent interest in theoretical physics, numerous results have contributed to its increase in popularity as a research topic. While many prior developments were incredibly important, the discovery that black holes radiate [86, 84] by Stephen Hawking in 1974 and the subsequent information paradox [85] he pointed out are clearly especially noteworthy in the expansion of the field.

Naturally, there are also many approaches to quantum gravity. Among them, string theory has produced remarkable results guiding us to our current understanding of gravity. One of the topics central to the work in this thesis is the idea of holography. The notion that a quantum description of a $d+1$ dimensional spacetime comes from a d dimensional theory also has its origins in through black hole thermodynamic [22, 86]. The discovery that the black hole entropy scales with the area, and not volume already indicated that the degrees of freedom responsible for describing them are localized to their 'boundary'. This idea was further developed in [173, 169]. In 1997, however, Maldacena's result provided an explicit string theory construction of this idea, the Anti-de Sitter/Conformal Field Theory correspondence [122].

AdS/CFT has been central to quantum gravity research since its conception, and provided some remarkable insights into how gravity can be described holographically. Despite the significant progress made over the past two decades, the dictionary relating quantities in the AdS bulk and the boundary CFT is still incomplete. One natural entry is the relationship between bulk and boundary operators, critical to answering many basic questions for phenomena where gravitational effects become important.

The very nature of holography, however, suggests a difficulty with this issue. In quantum field theory, field operators are associated with spacetime points, however, in a holographic setting, such an operator should be represented by objects that are localized to the boundary. This reconstruction process (the representation of bulk operators in terms of boundary

quantities) is now known to me non-unique and very non-local. This means that a bulk field has many distinct representations in terms of non-local boundary operators. This non-uniqueness of bulk reconstruction can be most naturally understood through the language of quantum error correction, and this will be further elaborated in chapter 7.

Often what is meant by the bulk reconstruction process is the representation of bulk fields built on a semiclassical background in terms of boundary quantities. Such a construction, however, assumes a priori knowledge of the bulk geometry. However, a complete entry to the dictionary would entail the ability to start exclusively with boundary information and obtain the information about the bulk metric as well as its semiclassical operators.

Nevertheless, obtaining the bulk metric is intimately related to understanding the emergence of the bulk in holography and the notion of locality holding (approximately) in gravity. While it is clear that locality should fail to hold near the planck scale, the equivalence principle, as well as experimental evidence, strongly supports the expectation from effective field theory that locality holds at low energy scales.

Despite this expectation, the recent emphasis on the inconsistencies at the heart of the black hole information paradox [10] suggest that understanding the emergence of locality and the applicability of local effective field theory will, at the very least, shine more light into one of the first, and still unresolved, questions in quantum gravity. The argument shows that the equivalence principle, local effective field theory, and unitarity are inconsistent in the black hole evaporation process. Studying this problem in the context of AdS/CFT has also failed to yield a satisfactory answer, in part due to the incomplete dictionary.

While AdS/CFT has given us a precise setting to study gravity, it is ultimately limited to spacetimes that are asymptotically anti-de Sitter. Part of the work presented in this thesis discusses a proposal that generalizes holography to general spacetimes. This work presented in chapter 5, based on the holographic screens initially presented by Raphael Bousso [28], shows that certain features present in AdS/CFT are more general than previously thought. The evidence presented for the holographic screen entanglement proposal is also at the heart of the emergence of locality in holography. The relationship between entanglement and geometry in the AdS/CFT correspondence is also largely believed to be critical to understanding the emergence of the bulk. The von Neumann entropy of a CFT subregion was proven to be equal to the area of a codimension two extremal surface anchored to the boundary to leading order proven in the context of the AdS/CFT correspondence. It turns out, this entangling surface is also extremely important in the reconstruction of bulk operators mentioned above.

It is clear that locality in gravity is intimately tied to numerous important phenomena and open questions. Studying the emergence of locality can be undertaken in many different ways, and such a multi-directional approach will likely be more fruitful in extracting the key mechanism underlying local physics as well as its failure in gravitational systems. With the exception of chapter 4, every topic contained in this thesis studies locality in gravity in through some lens.

The chapters in this thesis are organized in the following manner,

- The following chapter studies the observer dependence in quantum gravity and its relationship to the distribution of degrees of freedom associated with the quantum mechanical description of spacetime. This observer centered view proposed in [170] in the context of the information paradox, but shown to be insufficient. It is, nevertheless, intimately tied to the gauge redundancies present in quantum gravity.

- Chapter 3 presents a coherent picture for the black hole evaporation process that avoids the contradictions outlined in [10].

- In chapter 4, I present work related to the implications of large scale inflation and potential cosmological observations thereof to axions. In particular, it presents a consistent model that effectively suppresses otherwise large isocurvature fluctuations for axions in inflationary scenarios

- The holographic screen entanglement proposal discussed above is presented in chapter 5. There, the past or future holographic screens are used to present strong evidence for the proposal. In particular, it proves that the area of codimension two extremal surfaces anchored to subregions on past or future holographic screens satisfies the same inequalities as von Neumann entropy respects.

- Based on the holographic entanglement proposal presented, chapter 6 then studies its implications to the Hilbert space structure of such holographic theories for general spacetimes. In particular, it also studies the popular speculation of whether spacetime can be constructed from entanglement.

- The work presented in chapter 7 studies true reconstruction of the bulk, starting with purely boundary information. Ultimately, it presents a theorem which addresses what a local operator looks like within the context of the AdS/CFT correspondence. In doing so, the conformal metric for the spacetime is also obtained.

Chapter 2

Relativeness in Quantum Gravity: Limitations and Frame Dependence of Semiclassical Descriptions

2.1 Introduction

In the past decades, it has become increasingly apparent that the concept of spacetime must receive substantial revisions when it is treated in a fully quantum mechanical manner. The first clear sign of this came from the study of black hole physics [154]. Consider describing a process in which an object falls into a black hole, which eventually evaporates, from the viewpoint of a distant observer. Unitarity of quantum mechanics suggests that the information content of the object will first be stored in the black hole system, and then emitted back to distant space in the form of Hawking radiation [174, 165]. On the other hand, the equivalence principle implies that the object should not find anything special at the horizon, when the process is described by an observer falling with the object. These two pictures lead to inconsistency if we adopt the standard formulation of quantum field theory on curved spacetime, since it allows us to employ a class of equal time hypersurfaces (called nice slices) that pass through both the fallen object and late Hawking radiation, leading to violation of the no-cloning theorem of quantum mechanics [191].

In the early 90's, a remarkable suggestion to avoid this difficulty—called complementarity—was made [170]: the apparent cloning of the information occurring in black hole physics implies that the internal spacetime and horizon/Hawking radiation degrees of freedom appearing in different, i.e. infalling and distant, descriptions are not independent. This signals a breakdown of the naive global spacetime picture of general relativity at the quantum level, and it forces us to develop a new view of how classical spacetime arises in the full theory of quantum gravity. One of the main purposes of this paper is to present a coherent picture of this issue. We discuss how a series of well-motivated hypotheses leads to a consistent view of the effective emergence of global spacetime from a fundamental theory of quantum gravity.

In particular, we elucidate how this picture avoids the recently raised firewall paradox [10, 9, 126], which can be viewed as a refined version of the old information paradox [85]. Our analysis provides a concrete answer to how the information can be preserved at the quantum level in the black hole formation and evaporation processes.

A key element in developing our picture is to identify the origin and nature of the "entropy of spacetime," first discovered by Bekenstein and Hawking in studying black hole physics [22, 86]. In a previous paper [141], two of us argued that this entropy—the Bekenstein-Hawking entropy—is associated with the degrees of freedom that are coarse-grained *to obtain* the semiclassical description of the system: quantum theory of matter and radiation on a fixed spacetime background. This picture is consonant with the fact that in quantum mechanics, having a well-defined geometry of spacetime, e.g. a black hole in a well-defined spacetime location, requires taking a superposition of an enormous number of energy-momentum eigenstates, so we expect that there are many different ways to arrive at the same background for the semiclassical theory within the precision allowed by quantum mechanics. This implies that, when a system with a black hole is described in a distant reference frame, the information about the microstate of the black hole is *delocalized* over a large spatial region, since it is encoded globally in the way of taking the energy-momentum superposition to arrive at the geometry under consideration. In particular, we may naturally identify the spatial distribution of this information as that of the gravitational thermal entropy calculated using the semiclassical theory. This leads to a fascinating picture: the degrees of freedom represented by the Bekenstein-Hawking entropy play dual roles of spacetime and matter—they represent how the semiclassical geometry is obtained at the microscopic level and at the same time can be viewed as the origin of the thermal entropy, which is traditionally associated with thermal radiation in the semiclassical theory.

The delocalization of the microscopic information described above plays an important role in addressing the firewall/information paradox. As described in a distant reference frame, a general black hole state is specified by the following three classes of indices at the microscopic level:

- Indices labeling the (field or string theoretic) degrees of freedom in the exterior spacetime region, excited over the vacuum of the semiclassical theory;[1]
- Indices labeling the excitations of the stretched horizon;[2]
- Indices representing the degrees of freedom that are coarse-grained to obtain the semiclassical description, which we will collectively denote by k. The information in k represents how the black hole geometry is obtained at the microscopic level, and cannot be resolved by semiclassical operators. It is regarded as being delocalized following

[1]Note that the concepts of the breakdown of a semiclassical description and that of semiclassical *field* theory are not the same—there can be phase space regions in which an object can be well described as a string (or brane) propagating in spacetime, but not as a particle.

[2]The stretched horizon is located at a microscopic distance outside of the mathematical horizon, and is regarded as a physical (timelike) membrane which may be physically excited [170].

the spatial distribution of the gravitational thermal entropy, calculated using the semiclassical theory.

In a distant reference frame, an object falling into the black hole is initially described by the first class of indices, and then by the second when it hits the stretched horizon. The information about the fallen object will then reside there for, at least, time of order $Ml_{\rm P}^2 \ln(Ml_{\rm P})$ (the scrambling time [88, 161]), after which it will be transmitted to the index k. Here, M and $l_{\rm P}$ are the mass of the black hole and the Planck length, respectively. Finally, the information in k, which is delocalized in the whole zone region, will leave the black hole system through the Hawking emission, or black hole mining, process.

Since the microscopic information about the black hole is considered to be delocalized from the semiclassical standpoint, the Hawking emission, or black hole mining, process can be viewed as occurring at a *macroscopic* distance away from the stretched horizon without contradicting information conservation. In this region, degrees of freedom represented by the index k are converted into modes that have clear identities as semiclassical excitations, i.e. matter or radiation, above the spacetime background. This conversion process, i.e. the emission of Hawking quanta or the excitation of a mining apparatus, is accompanied by the appearance of negative energy excitations, which have *negative entropies* and propagate inward to the stretched horizon. As we will see, the microscopic dynamics of quantum gravity allows these processes to occur unitarily without violating causality among events described in low energy quantum field theory. This picture avoids firewalls as well as information cloning.

In the description based on a distant reference frame, a falling object can be described by the semiclassical theory only until it hits the stretched horizon, after which it goes outside the applicability domain of the theory. We may, however, describe the fate of the object using the semiclassical language somewhat longer by performing a *reference frame change*, specifically until the object hits a singularity, after which there is no reference frame that admits a semiclassical description of the object. This reference frame change is the heart of complementarity: the emergence of global spacetime in the classical limit. We argue that while descriptions in different reference frames (the descriptions before and after a complementarity transformation) apparently look very different, e.g. in locations of the degrees of freedom representing the microscopic information of the black hole, their predictions about the same physical question are consistent with each other. This consistency is ensured by an intricate interplay between the properties of microscopic information and the causal structure of spacetime.

It is striking that the concept of spacetime, e.g. the region in which a semiclassical description is applicable, depends on a reference frame. This extreme "relativeness" of the description is a result of nonzero Newton's constant $G_{\rm N}$. The situation is analogous to what happened when the speed of light, c, was realized to be finite [133]: in Galilean physics ($c = \infty$) a change of the reference frame leads only to a constant shift of all the velocities, while in special relativity ($c =$ finite) it also alters temporal and spatial lengths (time dilation and Lorentz contraction) and makes the concept of simultaneity relative. With gravity

($G_{\rm N} \neq 0$), even the concept of spacetime becomes relative. The trend is consistent—as we "turn on" fundamental constants in nature ($c = \infty \to$ finite and $G_{\rm N} = 0 \to \neq 0$), physical descriptions become more and more relative: descriptions of the same physical system in different reference frames appear to differ more and more.

The organization of this paper is the following. In Section 2.2, we discuss some basic aspects of the breakdown of global spacetime, setting up the stage for later discussions. In Sections 2.3 and 2.4, we describe how our picture addresses the problem of black hole formation and evaporation. We discuss the quantum structure of black hole microstates and the unitary flow of information as viewed from a distant reference frame (in Section 2.3), and how it can be consistent with the existence of interior spacetime (in Section 2.4). In particular, we elucidate how this picture addresses the arguments for firewalls and provides a consistent resolution to the black hole information paradox. In Section 2.5, we give our summary by presenting a grand picture of the structure of quantum gravity implied by our analysis of a system with a black hole.

Throughout the paper, we adopt the Schrödinger picture for quantum evolution, and use natural units in which $\hbar = c = 1$ unless otherwise stated. We limit our discussions to 4-dimensional spacetime, although we do not expect difficulty in extending to other dimensions. The value of the Planck length in our universe is $l_{\rm P} = G_{\rm N}^{1/2} \simeq 1.62 \times 10^{-35}$ m. A concise summary of the implications of our framework for black hole physics can be found in Ref. [136].

2.2 Failure of Global Spacetime

As described in the introduction, semiclassical theory applied to an entire global spacetime leads to overcounting of the true degrees of freedom at the quantum level. This implies that in the full theory of quantum gravity, a semiclassical description of physics emerges only in some limited sense. Here we discuss basic aspects of this limitation, setting up the stage for later discussions.

The idea of complementarity [170] is that the overcounting inherent in the global spacetime picture may be avoided if we limit our description to what a single "observer"—represented by a single worldline in spacetime—can causally access. Depending on which observer we choose, we obtain different descriptions of the system, which are supposed to be equivalent. Since the events an observer can see lie within the causal patch associated with the worldline representing the observer, we may assume that this causal patch is the spacetime region a single such description may represent. In particular, one may postulate the following [133, 132]:

- For a single description allowing a semiclassical interpretation of the system, the spacetime region represented is restricted to the causal patch associated with a single worldline. With this restriction, the description can be local in the sense that any physical correlations between low energy field theoretic degrees of freedom respect causality in

spacetime (beyond some microscopic quantum gravitational distance l_*, meaning that possible nonlocal corrections are exponentially suppressed $\sim e^{-r/l_*}$).

Depending on the worldline we take, we may obtain different descriptions of the same system, which are all local in appropriate spacetime regions. A transformation between different descriptions is nothing but the complementarity transformation.

To implement Hamiltonian quantum mechanics, we must introduce a time variable. This corresponds to foliating the causal patch by equal-time hypersurfaces, with a state vector $|\Psi(t)\rangle$ representing the state of the system on each hypersurface.[3] Let $\mathbf{x}$ be spatial coordinates parameterizing each equal-time hypersurface. Physical quantities associated with field theoretic degrees of freedom can then be obtained using field theoretic operators $\phi(\mathbf{x})$ and the state $|\Psi(t)\rangle$. (Excited string degrees of freedom will require the corresponding operators.) In general, the *procedure* of electing coordinates $(t, \mathbf{x})$, which we need to *define* states and operators, must be given independently of the background spacetime, since we do not know it a priori (and states may even represent superpositions of very different semiclassical geometries); an example of such procedures is described in Ref. [140]. In our discussions in this paper, however, we mostly consider issues addressed on a fixed background spacetime (at least approximately), so we need not be concerned with this problem too much—we may simply use any coordinate system adapted to a particular spacetime we consider, e.g. Schwarzschild-like coordinates for a black hole.

In the next two sections, we discuss how the complementarity picture described above works for a dynamical black hole. We discuss the semiclassical descriptions of the system in various reference frames, as well as their mutual consistency. In these discussions, we focus on a black hole that is well approximated by a Schwarzschild black hole in asymptotically flat spacetime. We do not expect difficulty in extending it to more general cases.

2.3 Black Hole—A Distant Description

Suppose we describe the formation and evaporation of a black hole in a distant reference frame. Following Ref. [174, 165], we postulate that there exists a unitary description which involves only the degrees of freedom that can be viewed as being on and outside the (stretched) horizon. To describe quantum states with a black hole, we adopt Schwarzschild-like time slicings to define equal-time hypersurfaces.[4] We argue that the origin of the Bekenstein-Hawking

[3] In general, the "time variable" of (constrained) Hamiltonian quantum mechanics may not be related directly with time we observe in nature [48]. Indeed, the whole "multiverse" may be represented by a state that does not depend on the time variable and is normalizable in an appropriate sense [134]. Even if this is the case, however, when we describe only a branch of the whole state, e.g. when we describe a system seen by a particular observer, the state *of the system* may depend on time. In this paper, we discuss systems with black holes, which are *parts of* the multiverse so their states may depend on time.

[4] Strictly speaking, to describe a general gravitating system we need a procedure to foliate the relevant spacetime region in a background independent manner, as discussed in the previous section. For our present purposes, however, it suffices to employ any foliation that reduces to Schwarzschild-like time slicings when

entropy may be viewed as a coarse-graining performed to obtain a semiclassical description of the evolving black hole. We then discuss implications of such a coarse-graining, in particular how it reconciles unitarity of the Hawking emission and black hole mining processes in the fundamental theory with the non-unitary (thermal) view in the semiclassical description.

Microscopic structure of a dynamical black hole

Consider a quantum state which represents a black hole of mass M located at some place at rest, where the position and velocity are measured with respect to some distant reference frame, e.g. an inertial frame elected at asymptotic infinity. Because of the uncertainty principle, such a state must involve a superposition of energy and momentum eigenstates. Let us first estimate the required size of the spread of energy ΔE, with E measured in the asymptotic region. According to the standard Hawking calculation, a state of a black hole of mass M will evolve after Schwarzschild time $\Delta t \approx O(Ml_{\rm P}^2)$ into a state representing a Hawking quantum of energy $\approx O(1/Ml_{\rm P}^2)$ and a black hole with the correspondingly smaller mass. The fact that these two states—before and after the emission—are nearly orthogonal implies that the original state must involve a superposition of energy eigenstates with

$$\Delta E \approx \frac{1}{\Delta t} \approx O\left(\frac{1}{Ml_{\rm P}^2}\right). \tag{2.1}$$

Of course, this is nothing but the standard time-energy uncertainty relation, and here we have assumed that a state after time $t \ll Ml_{\rm P}^2$ is not clearly distinguishable from the original one, so that the uncertainty relation is almost saturated.

Next, we consider the spread of momentum Δp, where p is again measured in the asymptotic region. Suppose we want to identify the spatial location of the black hole with precision comparable to the quantum stretching of the horizon $\Delta r \approx O(1/M)$, i.e. $\Delta d \approx O(l_{\rm P})$, where r and d are the Schwarzschild radial coordinate and the proper length, respectively. This implies that the superposition must involve momenta with spread $\Delta p \approx (1/Ml_{\rm P})(1/\Delta d) \approx O(1/Ml_{\rm P}^2)$, where the factor $1/Ml_{\rm P}$ in the middle expression is the redshift factor. This value of Δp corresponds to an uncertainty of the kinetic energy $\Delta E_{\rm kin} \approx p\Delta p/M \approx O(1/M^3 l_{\rm P}^4)$, which is much smaller than ΔE in Eq. (3.1). The spread of energy thus comes mostly from a superposition of different rest masses: $\Delta E \approx \Delta M$.

How many different independent ways are there to superpose the energy eigenstates to arrive at the same black hole geometry, at a fixed position within the precision specified by

the black hole exists. Note that macroscopic uncertainties in the black hole mass, location, and spin caused by the stochastic nature of Hawking radiation [146, 139] require us to focus on appropriate branches in the full quantum state in which the black hole in a given time has well-defined values for these quantities at the classical level. The relation between the Schwarzschild-like foliation and a general background independent foliation is then given by the standard coordinate transformation, which does not introduce subtleties beyond those discussed in this paper. The effect on unitarity by focusing on particular branches in this way is also minor, so we ignore it. The full unitarity, however, can be recovered by keeping all the branches in which the black hole has different classical properties at late times [139].

Δr and of mass M within an uncertainty of ΔM? We assume that the Bekenstein-Hawking entropy, $\mathcal{A}/4l_{\rm P}^2$, gives the logarithm of this number (at the leading order in expansion in inverse powers of $\mathcal{A}/l_{\rm P}^2$), where $\mathcal{A} = 16\pi M^2 l_{\rm P}^4$ is the area of the horizon. While the definition of the Bekenstein-Hawking entropy does not depend on the precise values of ΔM or Δp, a natural choice for these quantities is

$$\Delta M \approx \Delta p \approx O\left(\frac{1}{M l_{\rm P}^2}\right), \tag{2.2}$$

which we will adopt. The nonzero Bekenstein-Hawking entropy thus implies that there are exponentially many independent states in a small energy interval of $\Delta E \approx O(1/M l_{\rm P}^2)$. We stress that it is not appropriate to interpret this to mean that quantum mechanics introduces exponentially large degeneracies that do not exist in classical black holes. In classical general relativity, a set of Schwarzschild black holes located at some place at rest are parameterized by a continuous mass parameter M; i.e., there are a continuously infinite number of black hole states in the energy interval between M and $M + \Delta M$ for any M and small ΔM. Quantum mechanics *reduces* this to a finite number $\approx e^{S_0}\Delta M/M$, with S_0 given by[5]

$$S_0 = \frac{\mathcal{A}}{4l_{\rm P}^2} + O\left(\frac{\mathcal{A}^q}{l_{\rm P}^{2q}};\, q < 1\right). \tag{2.3}$$

This can also be seen from the fact that S_0 is written as $\mathcal{A}c^3/4l_{\rm P}^2\hbar$ when $\hbar$ and c are restored, which becomes infinite for $\hbar \to 0$.

As is clear from the argument above, there are exponentially many independent microstates, corresponding to Eq. (3.2), which are all black hole *vacuum* states: the states that do not have a field or string theoretic excitation on the semiclassical black hole background and in which the stretched horizon, located at $r_{\rm s} = 2Ml_{\rm P}^2 + O(1/M)$, is not excited.[6] Denoting the indices representing these exponentially many states collectively by k, which we call the *vacuum index*, basis states for the general microstates of a black hole of mass M (within the uncertainty of ΔM) can be given by

$$|\Psi_{\bar{a}\,a\,a_{\rm far};k}(M)\rangle. \tag{2.4}$$

Here, $\bar{a}$, a, and $a_{\rm far}$ represent the indices labeling the excitations of the stretched horizon, in the near exterior zone region (i.e. the region within the gravitational potential barrier

[5]Of course, quantum mechanics allows for a superposition of these finite number of independent states, so the number of possible (not necessarily independent) states is continuously infinite. The statement here applies to the number of independent states, regarding classical black holes with different M as independent states.

[6]These states can be defined, for example, as the states obtained by first forming a black hole of mass M and then waiting sufficiently long time after (artificially) switching off Hawking emission. Note that at the level of full quantum gravity, all the black hole states are obtained as excited states. Any semiclassical description, however, treats some of them as vacuum states on the black hole background.

defined, e.g., as $r \leq R_{\rm Z} \equiv 3Ml_{\rm P}^2$), and outside the zone ($r > R_{\rm Z}$), respectively.[7] As we have argued, the index k runs over $1, \cdots, e^{S_0}$ for the vacuum states $\bar{a} = a = a_{\rm far} = 0$. In general, the range for k may depend on $\bar{a}$ and a, but its dependence is higher order in $l_{\rm P}^2/\mathcal{A}$; i.e., for fixed $\bar{a}$ and a

$$k = 1, \cdots, e^{S_{\bar{a}a}}; \qquad S_{\bar{a}a} - S_0 \approx O\left(\frac{\mathcal{A}^q}{l_{\rm P}^{2q}};\, q < 1\right). \tag{2.5}$$

We thus mostly ignore this small dependence of the range of k on $(\bar{a}, a)$, i.e. the non-factorizable nature of the Hilbert space factors spanned by these indices, except when we discuss negative energy excitations associated with Hawking emission later, where this aspect plays a relevant role in addressing one of the firewall arguments.

Since we are mostly interested in physics associated with the black hole region, we also introduce the notation in which the excitations in the far exterior region are separated. As we will see later, the degrees of freedom represented by k can be regarded as being mostly in the region $r \leq R_{\rm Z}$, so we may write the states of the entire system in Eq. (3.3) as

$$|\Psi_{\bar{a}\, a\, a_{\rm far};k}(M)\rangle \approx |\psi_{\bar{a}a;k}(M)\rangle |\phi_{a_{\rm far}}(M)\rangle, \tag{2.6}$$

and call $|\psi_{\bar{a}a;k}(M)\rangle$ and $|\phi_{a_{\rm far}}(M)\rangle$ as the black hole and exterior states, respectively. Note that by labeling the states in terms of localized excitations, we need not write explicitly the trivial vacuum entanglement between the black hole and exterior states that does not depend on k, which typically exist when they are specified in terms of the occupation numbers of modes spanning the entire space.

How many independent quantum states can the black hole region support? Let us label appropriately coarse-grained excitations in the region $r_{\rm s} \leq r \leq R_{\rm Z}$ by $i = 1, 2, \cdots$, each of which carries entropy S_i. Suppose there are n_i excitations of type i at some fixed locations. The entropy of such a configuration is given by the sum of the "entropy of vacuum" in Eq. (3.2) and the entropies associated with the excitations:

$$S_I = S_0 + \sum_i n_i S_i. \tag{2.7}$$

The energy of the system in the region $r \leq R_{\rm Z}$ is given by the sum of the mass M of the black hole, which we define as the energy the system would have in the absence of an excitation outside the stretched horizon, and the energies associated with the excitations in the zone. Note that excitations here are defined as fluctuations with respect to a fixed background, so their energies E_i as well as entropies S_i can be either positive or negative, although the signs

[7] Strictly speaking, the states may also have the vacuum index associated with the ambient space in which the black hole exists. The information in this index, however, is not extracted in the Hawking evaporation or black hole mining process, so we ignore it here. (For more discussions, see, e.g., Section 5 of Ref. [141].) We will also treat excitations spreading both in the $r \leq R_{\rm Z}$ and $r > R_{\rm Z}$ regions only approximately by including them either in a or $a_{\rm far}$. The precise description of these excitations will require more elaborate expressions, e.g. than the one in Eq. (2.6), which we believe is an inessential technical subtlety in addressing our problem.

of the energy and entropy must be the same: $E_i S_i > 0$. The meaning of negative entropies will be discussed in detail in Sections 2.3 and 2.3.

Since excitations in the zone affect geometry, spacetime outside the stretched horizon, when they exist, is not exactly that of a Schwarzschild black hole. We require that these excitations do not form a black hole by themselves or become a part of the black hole at the center; otherwise, the state must be viewed as being built on a different semiclassical vacuum.[8] The total entropy S of the region $r \leq R_{\rm Z}$, i.e. the number of independent microscopic quantum states representing this region, is then given by

$$S = \ln\Big(\sum_I e^{S_I}\Big), \tag{2.8}$$

where I represents possible configurations of excitations, specified by the set of numbers $\{n_i\}$ and the locations of excitations of each type i, that do not modify the semiclassical vacuum in the sense described above. As suggested by a representative estimate [173], and particularly emphasized in Ref. [142], the contribution of such excitations to the total entropy is subdominant in the expansion in inverse powers of $\mathcal{A}/l_{\rm P}^2$: $S = S_0 + O(\mathcal{A}^q/l_{\rm P}^{2q}; q < 1)$. The total entropy in the near black hole region, $r \leq R_{\rm Z}$, is thus given by

$$S = \frac{\mathcal{A}}{4l_{\rm P}^2}, \tag{2.9}$$

at the leading order in $l_{\rm P}^2/\mathcal{A}$.

Emergence of the semiclassical picture and coarse-graining

The fact that all the independent microstates with different values of k lead to the same geometry suggests that the semiclassical picture is obtained after coarse-graining the degrees of freedom represented by this index; namely, any result in semiclassical theory is a statement about the maximally mixed ensemble of microscopic quantum states consistent with the specified background within the precision allowed by quantum mechanics [141]. According to this picture, the black hole vacuum state in the semiclassical description is given by the density matrix

$$\rho_0(M) = \frac{1}{e^{S_0}} \sum_{k=1}^{e^{S_0}} |\Psi_{\bar{a}=a=a_{\rm far}=0;k}(M)\rangle\langle\Psi_{\bar{a}=a=a_{\rm far}=0;k}(M)|. \tag{2.10}$$

Because of the coarse-graining of an enormous number of degrees of freedom, this density matrix has statistical characteristics.

[8] More precisely, we regard two geometries as being built on different classes of semiclassical vacua when they have different horizon configurations as viewed from a fixed reference frame. On the other hand, if two geometries have the same horizon, they belong to the same "vacuum equivalence class" in the sense that one can be converted into the other with "excitations." For more discussions on this point, see Ref. [140] and Section 2.3.

In order to obtain the response of this state to the operators in the semiclassical theory, we may trace out the subsystem on which they do not act. As we will discuss more later, the operators in the semiclassical theory in general act on a part, but not all, of the degrees of freedom represented by the k index. Let us denote the subsystem on which semiclassical operators act nontrivially by C, and its complement by $\bar{C}$. The index k may then be viewed as labeling the states in the combined $C\bar{C}$ system which satisfy certain constraints, e.g. the total energy being M within ΔM. The density matrix representing the semiclassical vacuum state in the Hilbert space in which the semiclassical operators act nontrivially, C, is given by

$$\tilde{\rho}_0(M) = \mathrm{Tr}_{\bar{C}}\, \rho_0(M). \tag{2.11}$$

Consistently with our identification of the origin of the Bekenstein-Hawking entropy, we assume that this density matrix represents the thermal density matrix with temperature $T_{\mathrm{H}} = 1/8\pi M l_{\mathrm{P}}^2$ in the zone region (as measured at asymptotic infinity):

$$\tilde{\rho}_0(M) \approx \frac{1}{\mathrm{Tr}\, e^{-\beta H_{\mathrm{sc}}(M)}} e^{-\beta H_{\mathrm{sc}}(M)}; \qquad \beta = \begin{cases} \frac{1}{T_{\mathrm{H}}} & \text{for } r \leq R_{\mathrm{Z}}, \\ +\infty & \text{for } r > R_{\mathrm{Z}}, \end{cases} \tag{2.12}$$

where $H_{\mathrm{sc}}(M)$ is the Hamiltonian of the semiclassical theory in the distant reference frame, which is defined in the region $r \geq r_{\mathrm{s}}$ on the black hole background of mass M.[9] (The meaning of position-dependent β is that the expression $\beta H_{\mathrm{sc}}(M)$ should be interpreted as β times the Hamiltonian density integrated over space.) Note that this procedure of obtaining Eq. (3.6) from Eq. (3.4) can be viewed as an example of the standard procedure of obtaining the canonical ensemble of a system from the microcanonical ensemble of a larger (isolated) system that contains the system of interest. In fact, if the system traced out is larger than the system of interest, $\dim \bar{C} \gtrsim \dim C$, we expect to obtain the canonical ensemble in this manner (see Ref. [144] for a related discussion). Below, we drop the tilde from the density matrix in Eq. (3.6), as it represents the same state as the one in Eq. (3.4)—$\rho_0(M)$ must be interpreted to mean either the right-hand side of Eq. (3.4) or of Eq. (3.6), depending on the Hilbert space under consideration.

In semiclassical field theory, the density matrix of Eq. (3.6) is obtained as a reduced density matrix by tracing out the region within the horizon in the *unique* global black hole vacuum state. Our view is that this density matrix, in fact, is obtained from a mixed state of exponentially many pure states, arising from a coarse-graining performed in Eq. (3.4); the prescription in the semiclassical theory provides (merely) a useful way of obtaining the same density matrix, in a similar sense in which the thermofield double state was originally introduced [176]. We emphasize that the information in k is invisible in the semiclassical theory (despite the fact that it involves subsystem C) as it is already coarse-grained *to obtain* the theory; in particular, the dynamics of the degrees of freedom represented by k cannot

[9] The Hilbert space of the semiclassical theory for states which have a single black hole at a fixed location at rest may be decomposed as $\mathcal{H} = \oplus_M \mathcal{H}_M$, where $\mathcal{H}_M$ is the space spanned by the states in which there is a black hole of (appropriately coarse-grained) mass M. In this language, $H_{\mathrm{sc}}(M)$ is a part of the semiclassical Hamiltonian acting on the subspace $\mathcal{H}_M$.

be described in terms of the semiclassical Hamiltonian $H_{\rm sc}(M)$.[10] As we will see explicitly later, it is this inaccessibility of k that leads to the apparent violation of unitarity in the semiclassical calculation of the Hawking emission process [85]. Note that because $\rho_0(M)$ takes the form of the maximally mixed state in k, results in the semiclassical theory do not depend on the basis of the microscopic states chosen in this space.

A comment is in order. In connecting the expression in Eq. (3.4) to Eq. (3.6), we have (implicitly) assumed that $|\Psi_{\bar{a}=a=a_{\rm far}=0;k}(M)\rangle$ represent the black hole vacuum states in the limit that the effect from evaporation is (artificially) shut off.[11] With this definition of vacuum states, the evolution effect necessarily "excites" the states, making $a \neq 0$, as we will see more explicitly in Section 2.3. As a consequence, the density matrix for the semiclassical operators representing the evolving black hole deviates from Eq. (3.6) even without matter or radiation. (In the semiclassical picture, this is due to the fact that the effective gravitational potential is not truly confining, so that the state of the black hole is not completely stationary.) If one wants, one can redefine vacuum states to be these states: the states that do not have any matter or radiation excitation on the *evolving* black hole background—the original vacuum states are then obtained as excited states on the new vacuum states.[12] This redefinition is possible because the two semiclassical "vacua" represented by the two classes of microstates belong to the same "vacuum equivalence class" in the sense described in the last paragraph of Section 2.3; specifically, they possess the same horizon for the same black hole mass, as defined for the evaporating case in Ref. [19].

As was mentioned above, semiclassical operators, in particular those for modes in the zone, act nontrivially on both a and k indices of microstates $|\Psi_{\bar{a}\,a\,a_{\rm far};k}(M)\rangle$. This can be seen as follows. If the operators acted only on the a index, the maximal mixture in k space with $a = 0$, Eq. (3.4), would look like a pure state from the point of view of these operators, contradicting the thermal nature in Eq. (3.6). On the other hand, if the operators acted only on the k index, they would commute with the maximally mixed state in k space, again contradicting the thermal state. Since the thermal nature of Eq. (3.6) is prominent only for modes whose energies as measured in the asymptotic region are of order the Hawking temperature or smaller

$$\omega \lesssim T_{\rm H}, \tag{2.13}$$

[10]This does not mean that a device made out of semiclassical degrees of freedom cannot probe information in k. Since there are processes in the fundamental theory (i.e. Hawking evaporation and mining processes) in which information in k is transferred to that in semiclassical *excitations* (i.e. degrees of freedom represented by the a and $a_{\rm far}$ indices), information in k can be probed by degrees of freedom appearing in the semiclassical theory. It is simply that these information extraction processes cannot be described within the semiclassical theory, since it can make statements only about the ensemble in Eq. (3.4) and excitations built on it.

[11]This is analogous to the treatment of a meta-stable vacuum in usual quantum field theory. At the most fundamental level (or on a very long timescale), such a state must be viewed as a scattering state built on the true ground state of the system. In practice (or on a sufficiently short timescale), however, we regard it as a vacuum state, which is approximately the ground state of a theory in which the tunneling out of this state is artificially switched off, e.g. by making the relevant potential barriers infinitely high.

[12]In the standard language in semiclassical theory, the original vacuum states correspond essentially to the Hartle-Hawking vacuum [82], while the new ones (very roughly) to the Unruh vacuum [179].

i.e. whose energies as measured by local (approximately) static observers are of order or smaller than the blueshifted Hawking temperature $T_{\rm H}/\sqrt{1-2Ml_{\rm P}^2/r}$, this feature is significant only for such infrared modes—operators representing modes with $\omega \gg T_{\rm H}$ act essentially only on the a index. For operators representing the modes with Eq. (3.8), their actions on microstates can be very complicated, although they act on the coarse-grained vacuum state of Eq. (3.4) as if it is the thermal state in Eq. (3.6), up to corrections suppressed by the exponential of the vacuum entropy S_0. The commutation relations of these operators defined on the coarse-grained states take the form as in the semiclassical theory, again up to exponentially suppressed corrections.

There is a simple physical picture for this phenomenon of "non-decoupling" of the a and k indices for the infrared modes. As viewed from a distant reference frame, these modes are "too soft" to be resolved clearly above the background—since the derivation of the semiclassical theory involves coarse-graining over microstates in which the energy stored in the region $r \lesssim R_{\rm Z}$ has spreads of order $\Delta E \approx 1/Ml_{\rm P}^2$, infrared modes with $\omega \lesssim T_{\rm H} \approx O(1/Ml_{\rm P}^2)$ are not necessarily distinguished from "spacetime fluctuations" of order ΔE. One might think that if a mode has nonzero angular momentum or charge, one can discriminate it from spacetime fluctuations. In this case, however, it cannot be clearly distinguished from vacuum fluctuations of a Kerr or Reissner-Nordström black hole having the corresponding (minuscule) angular momentum or charge. In fact, we may reverse the logic and view that this lack of a clear identity of the soft modes is the physical origin of the thermality of black holes (and thus of Hawking radiation).

Once the state for the vacuum of the semiclassical theory is obtained as in Eq. (3.4) (or Eq. (3.6) after partial tracing) and appropriate coarse-grained operators acting on it are identified, it is straightforward to construct the rest of the states in the theory—we simply have to act these operators (either field theoretic or of excited string states) on $\rho_0(M)$ to obtain the excited states. For example, to obtain a state which has a field theoretic excitation in the zone, one can apply the appropriate linear combination of creation and/or annihilation operators in the semiclassical theory, $a^\dagger_{\omega\ell m}$ and/or $a_{\omega\ell m}$:

$$\rho_{\bar{a}=0\,a\,a_{\rm far}=0}(M) = \left(\sum_{\ell,m}\int (c^a_{\omega\ell m}a_{\omega\ell m} + c'^a_{\omega\ell m}a^\dagger_{\omega\ell m})d\omega\right)\rho_0(M)\left(\sum_{\ell,m}\int (c^a_{\omega\ell m}a_{\omega\ell m} + c'^a_{\omega\ell m}a^\dagger_{\omega\ell m})d\omega\right)^\dagger, \tag{2.14}$$

where $c^a_{\omega\ell m}$ and $c'^a_{\omega\ell m}$ are coefficients. In the case that the applied operator is that for an infrared mode, this represents a state in which the thermal distribution for the infrared modes is "modulated" by an excitation over it. A construction similar to Eq. (2.14) also works for excitations in the far region. To obtain excitations of the stretched horizon, i.e. $\bar{a} \neq 0$, operators dedicated to describing them must be introduced. The detailed dynamics of these degrees of freedom, i.e. the $r = r_{\rm s}$ part of $H_{\rm sc}(M)$, is not yet fully known, however.

"Constituents of spacetime" and their distribution

While not visible in semiclassical theory, the black hole formation and evaporation (or mining) processes do involve the degrees of freedom represented by k, which we call *fine-grained vacuum degrees of freedom*, or vacuum degrees of freedom for short. The dynamics of these degrees of freedom as well as their interactions with the excitations in the semiclassical theory are determined by the fundamental theory of quantum gravity, which is not yet well known. We may, however, anticipate their basic properties based on some general considerations. In particular, motivated by the general idea of complementarity, we assume the following:

- Interactions with vacuum degrees of freedom do not introduce violation of causality among field theory degrees of freedom (except possibly for exponentially suppressed corrections, $\sim e^{-r/l_*}$ with l_* a short-distance quantum gravitational scale).
- Interactions between vacuum degrees of freedom and excitations in the semiclassical theory are such that unitarity is preserved at the microscopic level.

The first assumption is a special case of the postulate discussed in Section 2.2, applied to the distant reference frame description of a black hole. This implies that we cannot send superluminal signals among field theory degrees of freedom using interactions with vacuum degrees of freedom. The second assumption has an implication for how the vacuum degrees of freedom may appear from the semiclassical standpoint, which we now discuss.

In quantum mechanics, the information about a state is generally delocalized in space—locality is a property of dynamics, not that of states. In the case of black hole states, the information about k, which roughly represents slightly different "values" (superpositions) of M, is generally delocalized in a large spatial region, so that it can be accessed physically in a region away from the stretched horizon (e.g. around the edge of the zone $r \sim R_{\rm Z}$). This, however, does not mean that the complete information about the state can be recovered by a physical process occurring in a limited region in spacetime. For example, if we consider the set of e^{S_0} different black hole vacuum states, a physical detector occupying a finite spatial region can only partially discriminate these states in a given finite time.

To see how much information a physical detector in spatial region i can resolve, we can consider the reduced density matrix obtained after tracing out the subsystems that cannot be accessed by the semiclassical degrees of freedom associated with this region. In particular, we may consider the set of all field theory (and excited string state) operators that have support in i, and trace out the subsystems that do not respond to any of these operators (which we denote by $\bar{C}_i$):

$$\rho_0^{(i)} = \mathrm{Tr}_{\bar{C}_i}\, \rho_0(M), \tag{2.15}$$

where $\rho_0(M)$ is given by Eq. (3.4), and we have omitted the argument M for $\rho_0^{(i)}$. The von Neumann entropy of this density matrix, $S_0^{(i)} = -\mathrm{Tr}\, \rho_0^{(i)} \ln \rho_0^{(i)}$, then indicates the discriminatory power the region i possesses—a physical process occurring in region i can, at most, discriminate the e^{S_0} states into $e^{S_0^{(i)}}$ ($\ll e^{S_0}$) types in a characteristic timescale of

the system, $1/\Delta E \approx O(Ml_{\rm P}^2)$. According to the assumption in Eq. (3.6), this entropy is the gravitational thermal entropy contained in region i, calculated using the semiclassical theory.

We therefore arrive at the following picture. Let us divide the region $r \geq r_{\rm s}$ into N (arbitrary) subregions, each of which is assumed to have a sufficiently large number of degrees of freedom so that the thermodynamic limit can be applied. A basis state in the semiclassical theory can be written as

$$\rho_{\bar{a}\,a\,a_{\rm far}}(M) = \rho^{(1)}_{a_1} \otimes \rho^{(2)}_{a_2} \otimes \cdots \otimes \rho^{(N)}_{a_N}, \tag{2.16}$$

where $\rho^{(i)}_{a_i}$ are states defined in the i-th subregion, with a_i representing excitations contained in that region. (Following the convention in Section 2.3, we regard the vacuum states, $\bar{a} = a = a_{\rm far} = 0$, to be defined in the limit that the effect from evaporation is ignored.) Now, in the full Hilbert space of quantum gravity, there are e^{S_0} independent states that all reduce to the same $\rho_{\bar{a}\,a\,a_{\rm far}}(M)$ at the semiclassical level. These states can be written as

$$|\Psi_{\bar{a}\,a\,a_{\rm far};k=\{k_i\}}(M)\rangle = |\psi^{(1)}_{a_1;k_1}\rangle\,|\psi^{(2)}_{a_2;k_2}\rangle \cdots |\psi^{(N)}_{a_N;k_N}\rangle, \tag{2.17}$$

where $k_i = 1, \cdots, e^{S_0^{(i)}}$ with

$$S_0^{(i)} \approx \text{gravitational thermal entropy contained in subregion } i, \tag{2.18}$$

calculated using the semiclassical theory for subregions that do not contain the stretched horizon. The $S_0^{(i)}$'s for the subregions involving the stretched horizon are determined by the condition

$$\sum_{i=1}^{N} S_0^{(i)} = S_0 \approx \frac{\mathcal{A}}{4l_{\rm P}^2}, \tag{2.19}$$

which is valid in the thermodynamic limit. Assuming that the entropy on the stretched horizon is distributed uniformly on the surface, this condition determines the entropies contained in all the subregions.

The association of k_i's to each subregion, as in Eq. (2.17), corresponds to taking a specific basis in the space spanned by k. While the expressions above are strictly valid only in the thermodynamic limit, the corrections caused by deviating from it (e.g. due to correlations among subregions) do not affect our later discussions. In particular, it does not change the fact that the region around the edge of the zone, $r \leq R_{\rm Z}$ and $r - 2Ml_{\rm P}^2 \not\ll Ml_{\rm P}^2$, contains $O(1)$ bits of information about k (as it contains $O(1)$ bits of gravitational thermal entropy), which becomes important when we discuss the Hawking emission process in Section 2.3. Incidentally, the picture described here leads to the natural interpretation that the subsystem

that is traced out when going from Eq. (3.4) to Eq. (3.6) corresponds to the stretched horizon; i.e. $\bar{C}$ lives on the stretched horizon, while C in the zone.[13]

We stress that by the gravitational thermal entropy in Eq. (2.18), we mean that associated with the equilibrium vacuum state. It counts the thermal entropy within the zone, since this region is regarded as being in equilibrium because of its boundedness due to the stretched horizon and the potential barrier; on the other hand, Eq. (2.18) does not count the thermal entropy associated with Hawking radiation emitted from the zone, which is (artificially) switched off in defining our vacuum microstates. In other words, when calculating $S_0^{(i)}$'s using Eq. (2.18) we should use the vacuum state in Eq. (3.6), implying that we should use the local temperature, i.e. the temperature as measured by local static observers, of

$$T(r) \simeq \begin{cases} \frac{T_{\rm H}}{\sqrt{1-\frac{2Ml_{\rm P}^2}{r}}} & \text{for } r \leq R_{\rm Z}, \\ 0 & \text{for } r > R_{\rm Z}. \end{cases} \tag{2.20}$$

When the evolution effect is turned on, which we will analyze in Section 2.3, the state of the zone is modified ($a \neq 0$) due to an ingoing negative energy flux, while the state outside the zone is excited ($a_{\rm far} \neq 0$) by Hawking quanta, which are emitted from the edge of the zone and propagate freely in the ambient space. The contribution of the negative energy flux to the entropy within the zone is small, as we will see in Section 2.3.

The distribution of vacuum degrees of freedom in Eqs. (2.17, 2.18) is exactly the one needed for the interactions between these degrees of freedom and semiclassical excitations to preserve unitarity [141]. Imagine we put a physical detector at constant r in the zone. The detector then sees the thermal bath for all the modes with blueshifted Hawking temperature, Eq. (3.7), including higher angular momentum modes. This allows for the detector(s) to extract energy from the black hole at an accelerated rate compared with spontaneous Hawking emission: the mining process [180, 39]. In order for this process to preserve unitarity, the detector must also extract information at the correspondingly accelerated rate. This is possible if the information about the microstate of the black hole, specified by the index k, is distributed according to the gravitational thermal entropy, as in Eqs. (2.17, 2.18). A similar argument also applies to the spontaneous Hawking emission process, which is viewed as occurring around the edge of the zone, $r \sim R_{\rm Z}$, where the gravitational thermal entropy is small but not negligible. The microscopic and semiclassical descriptions of these processes will be discussed in detail in Sections 2.3 and 2.3.

It is natural to interpret the expression in Eq. (2.17) to mean that k_i labels possible configurations of "physical soft quanta"—or the "constituents of spacetime"—that comprise

[13] This in turn gives us a natural prescription to determine the location of the stretched horizon precisely. Since the semiclassical expression in Eq. (3.6) is expected to break down for $\ln \dim C > \ln \dim \bar{C}$, a natural place to locate the stretched horizon, i.e. the cutoff of the semiclassical spacetime, is where the gravitational thermal entropy outside the stretched horizon becomes $S_0/2 = \mathcal{A}/8l_{\rm P}^2$. For n low energy species, this yields $r_{\rm s} - 2Ml_{\rm P}^2 \sim n/M \sim l_*^2/Ml_{\rm P}^2$, where l_* is the string (cutoff) scale and we have used the relation $l_*^2 \sim nl_{\rm P}^2$, which is expected to apply in any consistent theory of quantum gravity (see, e.g., Ref. [55]). This scaling is indeed consistent, giving the local Hawking temperature at the stretched horizon $T(r_{\rm s}) \sim 1/l_*$, where $T(r)$ is given in Eq. (3.7).

the region i. In a certain sense, this interpretation is correct. The dimension of the relevant Hilbert space, $e^{S_0^{(i)}}$, controls possible interactions of the vacuum degrees of freedom with the excitations in the semiclassical theory in region i, e.g. how much information a detector located in region i can extract from the vacuum degrees of freedom. This simple picture, however, breaks down when we describe the same system from a different reference frame. As we will discuss in Section 2.4, the distribution of the vacuum degrees of freedom *depends on the reference frame*—they are not "anchored" to spacetime. Nevertheless, in a fixed reference frame, the concept of the spatial distribution of the degrees of freedom represented by the index k does make sense. In particular, in a distant reference frame the distribution is given by the gravitational thermal entropy calculated in the semiclassical theory, as we discussed here.

Hawking emission—"microscopic" and semiclassical descriptions

The formation and evaporation of a black hole involve processes in which the information about the initial collapsing matter is transferred into the vacuum index k, which will later be transferred back to the excitations in the semiclassical theory, i.e. the state of final Hawking radiation. Schematically, we may write these processes as

$$|m_{\text{init}}\rangle \;\to\; \sum_{k=1}^{e^{S_0(M(t))}} \sum_{l} c_{kl}(t)\, |\psi_k(M(t))\rangle\, |r_l(t)\rangle \;\to\; |r_{\text{fin}}\rangle, \tag{2.21}$$

where $|m_{\text{init}}\rangle$, $|\psi_k(M(t))\rangle$, $|r_l(t)\rangle$, and $|r_{\text{fin}}\rangle$ represent the states for the initial collapsing matter, the black hole of mass $M(t)$ (which includes the near exterior zone region; see Eq. (2.6)), the subsystem complement to the black hole at time t, and the final Hawking quanta after the black hole is completely evaporated, respectively. Here, we have suppressed the indices representing excitations for the black hole states. For generic initial states and microscopic emission dynamics, this evolution satisfies the behavior outlined in Ref. [145] on general grounds.

In this subsection, we discuss how the black hole evaporating process in Eq. (2.21) proceeds in details, elucidating how the arguments for firewalls in Refs. [10, 9, 126] are avoided. We also discuss how the semiclassical theory describes the same process, elucidating how the thermality of Hawking radiation arises despite the unitarity of the process at the fundamental level.

"Microscopic" (unitary) description

Let us first consider how the "elementary" Hawking emission process is described at the microscopic level,[14] i.e. how a "single" Hawking emission occurs in the absence of any exci-

[14]By the "microscopic" description, we mean a description in which the vacuum index k is kept (i.e. not coarse-grained as in the semiclassical description) so that the process is manifestly unitary at each stage of the evolution. A complete description of the microscopic dynamics of the vacuum degrees of freedom requires the fundamental theory of quantum gravity, which is beyond the scope of this paper.

tations other than those directly associated with the emission. (As we will see later, this is not a very good approximation in general, but the treatment here is sufficient to illustrate the basic mechanism by which the information is transferred from the black hole to the ambient space.)

Suppose a black hole of mass M is in microstate k:

$$|\Psi_k(M)\rangle = |\psi_k(M)\rangle|\phi_I\rangle, \tag{2.22}$$

where $|\psi_k(M)\rangle$ is the black hole state, in which we have omitted indices representing excitations, while $|\phi_I\rangle$ is the exterior state, from which we have suppressed small M dependence (which, e.g., causes a small gravitational redshift of a factor of about 1.5 for the emitted Hawking quanta to reach the asymptotic region). As discussed in Sections 2.3 and 2.3, we consider $|\Psi_k(M)\rangle$ to be one of the black hole vacuum microstates in the limit that the evolution effect is shut off; see, e.g., Eqs. (3.6, 3.7). The effect of the evolution, which consists of successive elementary Hawking emission processes, will be discussed later.

After a timescale of $t \approx O(Ml_{\rm P}^2)$, the state in Eq. (3.9) evolves due to Hawking emission as

$$|\psi_k(M)\rangle|\phi_I\rangle \to \sum_{i,a,k'} c^k_{iak'}|\psi_{a;k'}(M)\rangle|\phi_{I+i}\rangle, \tag{2.23}$$

where $|\phi_{I+i}\rangle$ is the state in which newly emitted Hawking quanta, labeled by i and having total energy E_i, are added to the appropriately time evolved $|\phi_I\rangle$. The index a represents the fact that the black hole state has negative energy excitations of total energy $-E_a$ ($E_a > 0$) around the edge of the zone, created in connection with the emitted Hawking quanta; the coefficients $c^k_{iak'}$ are nonzero only if $E_i \approx E_a$ (within the uncertainty).[15] The negative energy excitations then propagate inward, and after a time of order $Ml_{\rm P}^2 \ln(Ml_{\rm P})$ collide with the stretched horizon, making the black hole states relax as

$$|\psi_{a;k'}(M)\rangle \to \sum_{k_a} d^{ak'}_{k_a}|\psi_{k_a}(M - E_a)\rangle. \tag{2.24}$$

The combination of Eqs. (3.10, 3.11) yields

$$|\psi_k(M)\rangle|\phi_I\rangle \to \sum_{i,k_i} \alpha^k_{ik_i}|\psi_{k_i}(M - E_i)\rangle|\phi_{I+i}\rangle, \tag{2.25}$$

where $\alpha^k_{ik_i} = \sum_{a,k'} c^k_{iak'} d^{ak'}_{k_i}$, and we have used $E_i = E_a$; here, $M - E_i$ for different i may belong to the same mass within the precision ΔM, i.e. $M - E_i = M - E_{i'}$ for $i \neq i'$. This expression shows that information in the black hole can be transferred to the radiation state i.

[15] To be precise, the sum in the right-hand side of Eq. (3.10) contains the "$i = 0$ terms" representing the branches in which no quantum is emitted: $|\phi_{I+0}\rangle = |\phi_I\rangle$. In these terms, there is no negative energy excitation: $c^k_{0ak'} \neq 0$ only for $a = 0$. The following expressions are valid including these terms with the definition $E_{i=0} = E_{a=0} = 0$.

It is important that the negative energy excitations generated in Eq. (3.10) come with *negative entropies*, so that each of the processes in Eqs. (3.10, 3.11) (as well as the propagation of the negative energy excitations in the zone) is separately unitary. This means that as k and i run over all the possible values with a being fixed, the index k' runs only over $1, \cdots, e^{S_0(M-E_a)}$, the dimension of the space spanned by k_a. In fact, this is an example of the non-factorizable nature of the Hilbert space factors spanned by k and a discussed in Eq. (2.5), which we assume to arise from the fundamental theory. This structure of the Hilbert space allows for avoiding the argument for firewalls in Ref. [9]—unlike what is imagined there, elements of the naive Fock space built on each k in a way isomorphic to that of quantum field theory are not all physical; the physical Hilbert space is smaller than such a (hypothetical) Fock space. This implies, in particular, that the Fock space structure of a semiclassical theory does not factor from the space spanned by the vacuum index k, as is also implied by the analysis in Section 2.3.

To further elucidate the point made above, we can consider the following simplified version of the relevant processes. Suppose a black hole in a superposition state of $|\psi_k(M)\rangle$'s ($k = 1, \cdots, e^{S_0(M)}$) releases 1 bit of information through Hawking emission of the form:

$$|\psi_k(M)\rangle|\phi_0\rangle \to \begin{cases} |\psi_{a;\frac{k+1}{2}}(M)\rangle|\phi_1\rangle & \text{if } k \text{ is odd,} \\ |\psi_{a;\frac{k}{2}}(M)\rangle|\phi_2\rangle & \text{if } k \text{ is even,} \end{cases} \tag{2.26}$$

where we have assumed $E_1 = E_2 = (\ln 2)/8\pi M l_{\rm P}^2 \simeq T_{\rm H}$, so that the entropy of the black hole after the emission is reduced by 1 bit: $S_0(M - E_1) = S_0(M) - \ln 2$. Note that the index representing the negative energy excitation (of energy $-E_1$) takes the same value a in the first and second lines. Namely, while the entire process in Eq. (2.26) is unitary, the initial states with $k = 2n - 1$ and $2n$ lead to the *same black hole state*. After the negative energy excitation reaches the stretched horizon, the black hole states relax into vacuum states for a smaller black hole:

$$|\psi_{a;k'}(M)\rangle \to |\psi_{k_1=k'}(M - E_1)\rangle. \tag{2.27}$$

While the resulting black hole has a smaller entropy than the original black hole, this relaxation process is unitary because k' in the left-hand side runs only over $1, \cdots, e^{S_0(M)}/2 = e^{S_0(M-E_1)}$. We note that the creation of a positive energy Hawking quantum and a negative energy excitation in Eq. (2.26) (and in Eq. (3.10)) takes a form very different from the standard "pair creation" of particles, which is often invoked to visualize the Hawking emission process. In the pair creation picture, the positive and negative energy excitations are maximally entangled with each other, which is not the case here. In fact, it is this lack of entanglement that allows the emission process to transfer the information from the black hole to radiation.

We emphasize that from the semiclassical spacetime viewpoint, the emission of Eq. (3.10) is viewed as occurring locally around the edge of the zone, which is possible because the information about the black hole microstate extends into the whole zone region according

to Eqs. (2.17, 2.18). To elucidate this point, we may consider the tortoise coordinate

$$r^* = r + 2Ml_{\rm P}^2 \ln \frac{r - 2Ml_{\rm P}^2}{2Ml_{\rm P}^2}, \tag{2.28}$$

in which the region outside the Schwarzschild horizon $r \in (2Ml_{\rm P}^2, \infty)$ is mapped into $r^* \in (-\infty, \infty)$. This coordinate is useful in that the kinetic term of an appropriately redefined field takes the canonical form, so that its propagation can be analyzed as in flat space. In this coordinate, the stretched horizon, located at $r = 2Ml_{\rm P}^2 + O(l_*^2/Ml_{\rm P}^2)$ (see footnote 13), is at

$$r_{\rm s}^* \simeq -4Ml_{\rm P}^2 \ln \frac{Ml_{\rm P}^2}{l_*} \simeq -4Ml_{\rm P}^2 \ln(Ml_{\rm P}), \tag{2.29}$$

where l_* is the string (or gravitational cutoff) scale, which we take to be within a couple of orders of magnitude of $l_{\rm P}$. This implies that there is a large distance between the stretched horizon and the potential barrier region when measured in r^*: $\Delta r^* \approx 4Ml_{\rm P}^2 \ln(Ml_{\rm P}) \gg O(Ml_{\rm P}^2)$ for $\ln(Ml_{\rm P}) \gg 1$. On the other hand, a localized Hawking quantum is represented by a wavepacket with width of $O(Ml_{\rm P}^2)$ in r^*, since it has an energy of order $T_{\rm H} = 1/8\pi Ml_{\rm P}^2$ defined in the asymptotic region.

The point is that, given the state $|\Psi_k(M)\rangle = |\psi_k(M)\rangle|\phi_I\rangle$, the process in Eq. (3.10) occurs in the region $|r^*| \approx O(Ml_{\rm P}^2)$ (i.e. the region in which the effective gravitational potential starts shutting off toward large r^*) without involving deep interior of the zone $-r^* \gg Ml_{\rm P}^2$. In this region, information stored in the vacuum state is converted into that of a particle state outside the zone. More specifically, the information in the vacuum represented by the k index (which may also be viewed as a thermal bath of infrared modes, Eq. (3.8), though only in certain senses) is transferred into that in modes $a_{\rm far} \neq 0$, i.e. Hawking quanta, which have clear independent identities over the background spacetime. Due to energy conservation, this process is accompanied by the creation of ingoing negative energy excitations; however, they are not maximally entangled with the emitted Hawking quanta.

In Fig. 2.1, we depict schematically the elementary Hawking emission process described here. In the figure, we have denoted the emitted Hawking quanta as well as negative energy excitations by arrows, although they are mostly s-waves [148]. The discussion here makes it clear that the purifiers of the emitted Hawking quanta in the Hawking emission process are microstates which semiclassical theory describes as a vacuum. In particular, the emission process does *not* involve any excitation which, in the near horizon Rindler approximation, appears as a mode breaking entanglement between the two Rindler wedges necessary to keep the horizon smooth. Outgoing Hawking quanta emerge at the edge of the zone, living outside the applicability of the Rindler approximation. Ingoing negative energy excitations appear, in the Rindler approximation, as modes smooth in Minkowski space, which involve necessary entanglements between Rindler modes in the two wedges and have frequencies of order $1/Ml_{\rm P}^2$ in the Minkowski frame. Unlike what was considered in Ref. [10], and unlike what a "naive" interpretation of semiclassical theory might seem to suggest, Hawking quanta are not modes associated solely with one of the Rindler wedges (b modes in the notation

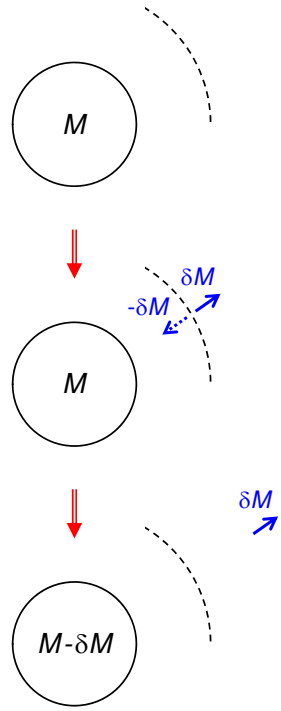
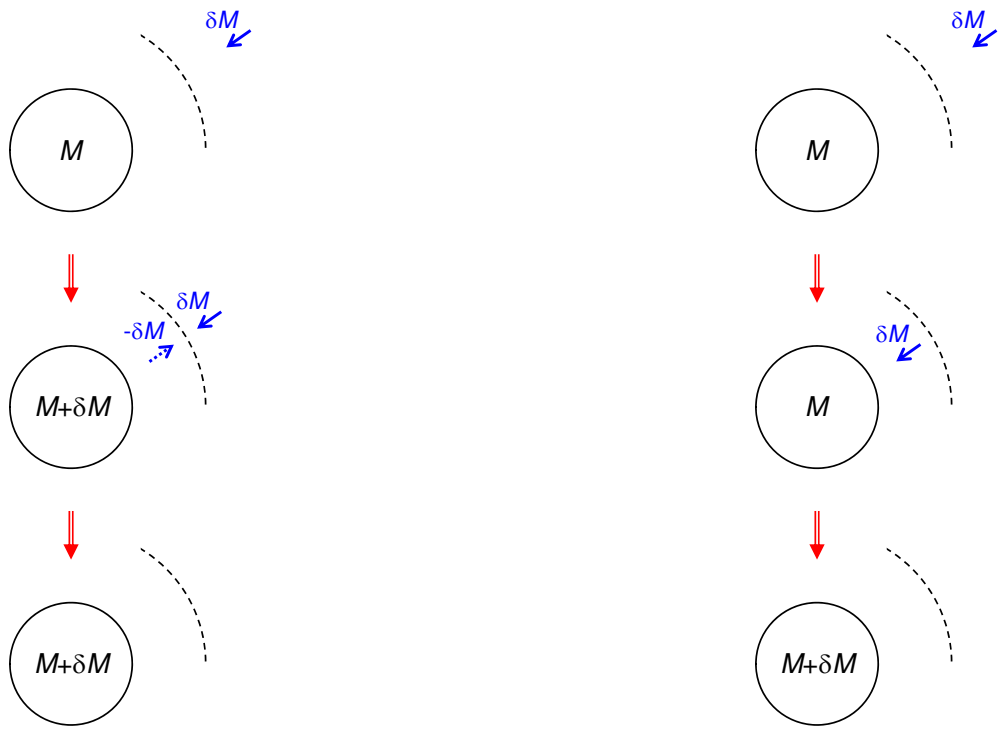


Figure 2.1: A schematic picture of the elementary Hawking emission process; time flows from the top to the bottom. The edge of the zone, i.e. the barrier region of the effective gravitational potential, is shown by a portion of a dashed circle at each moment in time. The emitted Hawking quanta as well as negative energy excitations are depicted by arrows (solid and dotted, respectively) although they are mostly s-waves.

of Ref. [10]) nor outgoing Minkowski modes (a modes), which would appear to have high energies for observers who are freely falling into the black hole. This allows for avoiding the entropy argument for firewalls given in Ref. [10] as well as the typicality argument in Ref. [126].

In the discussion of the Hawking emission so far, we have assumed that a single emission of Hawking quanta as well as the associated creation of ingoing negative energy excitations occur in a black hole vacuum state consisting of $|\Psi_k(M)\rangle$'s, which are defined in the limit that the evolution effect is ignored. In reality, however, there are always of order $\ln(Ml_{\rm P})$ much of negative energy excitations in the zone, since the emission process occurs in every time interval of order $Ml_{\rm P}^2$ and the time it takes for a negative energy excitation to reach the stretched horizon is of order $Ml_{\rm P}^2 \ln(Ml_{\rm P})$ (both measured in the asymptotic region)—an evaporating black hole has an ingoing flux of negative energy excitations of entropy $\approx O(-\ln(Ml_{\rm P}))$ at all times. This flux of excitations modifies spacetime geometry from that of a Schwarzschild black hole; in particular, the geometry near the horizon is well



(a) Time reversal of Fig. 2.1

(b) Generic incoming radiation

Figure 2.2: Time reversal of the Hawking emission process (a) as opposed to the process in which generic incoming radiation enters into the zone of a usual black hole (b). The former is an entropy decreasing process requiring an exponentially fine-tuned initial state, while the latter is a standard process respecting the (generalized) second law of thermodynamics.

described by the advanced/ingoing Vaidya metric [19]. Note that as discussed in Section 2.3, we may redefine our vacuum states to include these negative energy excitations, although we do not do it here.

Finally, it is instructive to consider the time reversal of the Hawking emission process. In this case, radiation coming from the far exterior region and outgoing negative energy excitations emitted from the stretched horizon meet around the edge of the zone; see Fig. 2.2(a). This results in a black hole state of mass given by the sum of the mass M of the original black hole (before emitting the negative energy excitations) and the energy δM of the incoming radiation. It is a "vacuum" state in the sense that there is no excitation in the zone except for those associated with a steady flux of outgoing negative energy excitations. We emphasize that this process is very different from what happens when generic incoming radiation of energy $\delta M \approx O(1/Ml_{\rm P}^2)$ is sent to a usual (i.e. evaporating, not anti-evaporating) black hole. In this case, the radiation enters into the zone without being "annihilated" by a negative energy excitation, which after hitting the stretched horizon will lead to a black hole

state of mass $M + \delta M$; see Fig. 2.2(b). In fact, the process in Fig. 2.2(a) is a process which leads to a decrease of coarse-grained (or thermal) entropy, as implied by the fact that the coarse-grained entropy increases in the standard Hawking emission process [194]. In order for this to happen, therefore, the initial radiation and black hole state must be exponentially fine-tuned; otherwise, the radiation would simply propagate inward in the zone as depicted in Fig. 2.2(b) (although it can be subject to significant scattering by the effective gravitational potential at the time of the entrance). The origin of the conversion from radiation to vacuum degrees of freedom for such a fine-tuned initial state can be traced to the non-decoupling of the a and k indices discussed in Section 2.3.[16]

Semiclassical (thermal) description

The expression in Eq. (2.21) implies that at an intermediate stage of the evolution, the information about the initial collapsing matter is encoded in the black hole microstates labeled by k and their entanglement with the rest of the system (which will later be transformed into the state of final-state Hawking radiation). Since semiclassical theory is incapable of describing the dynamics associated with the index k, it leads to apparent violation of unitarity *at all stages* of the black hole formation and evaporation processes. In particular, the state of the emitted Hawking quanta in each time interval of order $M(t)l_{\rm P}^2$ is given by the incoherent thermal superposition with temperature $1/8\pi M(t)l_{\rm P}^2$, making the final Hawking radiation state a mixed thermal state—this is an intrinsic limitation of the semiclassical description, which involves a coarse-graining.

To see in detail how thermal Hawking radiation in the semiclassical picture results from unitary evolution at the fundamental level, let us analyze the elementary Hawking emission process given in Eq. (3.12). Following Eq. (3.4), we consider the "semiclassical vacuum state" with a black hole of mass M, obtained after taking the maximally mixed ensemble of microstates:

$$\rho(M) = \frac{1}{e^{S_0(M)}} \sum_{k=1}^{e^{S_0(M)}} |\psi_k(M)\rangle|\phi_I\rangle\langle\psi_k(M)|\langle\phi_I|. \tag{2.30}$$

The evolution of this state under Eq. (3.12) is then given by

$$\rho(M) \to \frac{1}{e^{S_0(M)}} \sum_{k=1}^{e^{S_0(M)}} \sum_{i,i'} \sum_{k_i=1}^{e^{S_0(M-E_i)}} \sum_{k'_{i'}=1}^{e^{S_0(M-E_{i'})}} \alpha^k_{ik_i}\alpha^{k*}_{i'k'_{i'}} |\psi_{k_i}(M-E_i)\rangle|\phi_{I+i}\rangle\langle\psi_{k'_{i'}}(M-E_{i'})|\langle\phi_{I+i'}|. \tag{2.31}$$

[16] If the black hole vacuum states are redefined as discussed in Section 2.3, the outgoing negative energy flux cannot be seen as excitations. The physics described here, however, will not change; in particular, only exponentially fine-tuned initial states allow for converting radiation to vacuum degrees of freedom around the edge of the zone.

Now, assuming that the microscopic dynamics of the vacuum degrees of freedom are generic, we expect using $S_0(M) = 4\pi M^2 l_{\rm P}^2$ that tracing out the black hole states leads to

$$\mathrm{Tr}\left[\frac{1}{e^{S_0(M)}}\sum_{k=1}^{e^{S_0(M)}}\sum_{k_i=1}^{e^{S_0(M-E_i)}}\sum_{k'_{i'}=1}^{e^{S_0(M-E_{i'})}}\alpha^k_{ik_i}\alpha^{k*}_{i'k'_{i'}}|\psi_{k_i}(M-E_i)\rangle\langle\psi_{k'_{i'}}(M-E_{i'})|\right]\approx\frac{1}{Z}g_i e^{-\frac{E_i}{T_{\rm H}}}\delta_{ii'}, \tag{2.32}$$

where $T_{\rm H} = 1/8\pi M l_{\rm P}^2$, $Z = \sum_i g_i e^{-E_i/T_{\rm H}}$, and g_i is a factor that depends on i. This allows us to write the reduced density matrix representing the exterior state after the evolution in Eq. (2.31) as

$$\rho_{\rm ext}\approx\frac{1}{Z}\sum_i g_i e^{-\frac{E_i}{T_{\rm H}}}|\phi_{I+i}\rangle\langle\phi_{I+i}|, \tag{2.33}$$

which is the result obtained in Hawking's original calculation, with g_i representing the gray-body factor calculable in the semiclassical theory [148].

The analysis given above elucidates why the semiclassical calculation sees apparent violation of unitarity in the Hawking emission process, i.e. why the final expression in Eq. (2.33) does not depend on microstates of the black hole, despite the fact that the elementary process in Eq. (3.12) is unitary, so that the coefficients $\alpha^k_{ik_i}$ depend on k. It is because the semiclassical calculation (secretly) deals with the mixed state, Eq. (2.30), from the beginning—states in semiclassical theory are maximal mixtures of black hole microstates labeled by vacuum indices, i.e. k's. By construction, the semiclassical theory cannot capture unitarity of detailed microscopic processes involving these indices, including the black hole formation and evaporation processes.

We finally discuss how the unitarity and thermal nature of the black hole evaporation process may appear in (thought) experiments, illuminating physical implications of the picture described here. Suppose we prepare an ensemble of a large number of black holes of mass M all of which are in an identical microstate k, and collect the Hawking quanta emitted from these black holes in a time interval of order $M l_{\rm P}^2$. The quanta emitted from each black hole are then in the same quantum state throughout the ensemble, so that a measurement of the spectrum of all the emitted quanta does *not* reveal the thermal property predicted by the semiclassical theory. On the other hand, if the members of the ensemble are in different microstates distributed randomly in k space, then the collection of the Hawking quanta emitted from all the black holes do exhibit the thermal nature consistent with the prediction of the semiclassical theory within the Hilbert space describing the quanta emitted from *each* black hole (which has dimension only of order unity).

What is the significance of the thermal nature for a single black hole, rather than an ensemble of a large number of black holes? If we form a black hole of mass M in a particular microstate k and collect all the Hawking quanta emitted throughout the evaporation process *without measuring them along the way*, then the state of the quanta contains the complete information about k, reflecting unitarity of the process at the fundamental level—the concept of thermality does not apply to this particular state *as a whole*. On the other hand, if an observer measures Hawking quanta emitted in each time interval of order $M(t)l_{\rm P}^2$, then the

(incoherent) ensemble of measurement outcomes does exhibit the thermal nature as predicted by the semiclassical theory.[17] Since this is the kind of measurement that a realistic observer typically makes, the semiclassical theory can be said to provide a good prediction even for the outcome of (a series of) measurements a single observer performs on a single black hole.

Black hole mining—"microscopic" and semiclassical descriptions

It is known that one can accelerate the energy loss rate of a black hole faster than that of spontaneous Hawking emission by extracting its energy from the thermal atmosphere using a physical apparatus: the mining process. This acceleration occurs largely because the number of "channels" one can access increases by going into the zone—unlike the case of spontaneous Hawking emission, which is dominated by s-wave radiation, higher angular momentum modes can also contribute to the energy loss in this process [39]. Note that the rate of energy loss associated with *each* channel, however, is still the same order as that in the spontaneous Hawking emission process: energy of order $E \approx O(1/Ml_{\mathrm{P}}^2)$ is lost in each time interval of $t \approx O(Ml_{\mathrm{P}}^2)$, with E and t both defined in the asymptotic region. This fact will become important in Section 2.4 when we discuss the mining process as viewed from an infalling reference frame.

The information transfer associated with the mining process occurs in a similar way to that in the spontaneous Hawking emission process. An essential difference is that since the process involves higher angular momentum modes, the negative energy excitations arising from backreactions can now be localized in angular directions. Specifically, consider a physical detector (or a system of detectors) located at a fixed Schwarzschild radial coordinate $r = r_{\mathrm{d}}$ within the zone, $r_{\mathrm{s}} < r_{\mathrm{d}} < R_{\mathrm{Z}}$. The detector then responds as if it is immersed in the thermal bath of blueshifted Hawking temperature $T(r_{\mathrm{d}})$, with $T(r)$ given by Eq. (3.7). Suppose the detector has the ground state $|d_0\rangle$ and excited states $|d_i\rangle$ $(i = 1, 2, \dots)$ playing the role of the "ready" state and pointer states, respectively, and that the proper energies needed to excite $|d_0\rangle$ to $|d_i\rangle$ are given by $E_{\mathrm{d},i}$. The mining process can then be written such that after a timescale of $t \approx O(Ml_{\mathrm{P}}^2)$ (as measured in the asymptotic region), the state of the combined black hole and detector system evolves as

$$|\psi_k(M)\rangle|d_0\rangle \to \sum_{i,a,k'} c_{iak'}^k |\psi_{a;k'}(M)\rangle|d_i\rangle, \tag{2.34}$$

where we have assumed, as in the discussion of "elementary" Hawking emission, that there are no excitations other than those directly associated with the process. The state $|\psi_{a;k'}(M)\rangle$

[17] In the more fundamental, many-world picture, this implies that the record of a physical observer who has "measured," or interacted with, emitted quanta in multiple moments shows a result consistent with the thermality predicted by the semiclassical theory. Note that a single branch in which such an observer lives does *not* in general contain the whole information about the initial black hole state k. The complete information about k (as well as that of the initial state of the observer) is contained only in a state given by a superposition of all possible branches resulting from interactions (and non-interactions) between the observer and quanta, representing all the possible "outcomes" the observer could have had (the probability distribution of which is consistent with thermality).

arises as a result of backreaction of the detector response; it contains a negative energy excitation a with energy $-E_a$, which is generally localized in angular directions. The coefficients $c^k_{iak'}$ are nonzero only if $E_a \approx E_{\mathrm{d},i}\sqrt{1-2Ml_{\mathrm{P}}^2/r_{\mathrm{d}}}$ within the uncertainty.

Once created, the negative energy excitations propagate inward, and after time of $t \approx r^*_{\mathrm{d}} - r^*_{\mathrm{s}}$ collide with the stretched horizon, where r^* is the tortoise coordinate in Eq. (2.28). This will make the black hole states relax as

$$|\psi_{a;k'}(M)\rangle \to \sum_{k_a} d^{ak'}_{k_a} |\psi_{k_a}(M-E_a)\rangle, \tag{2.35}$$

in the scrambling time of $t \approx O(Ml_{\mathrm{P}}^2 \ln(Ml_{\mathrm{P}}))$. As in the case of spontaneous Hawking emission, this relaxation process is unitary because the negative energy excitations carry negative entropies; i.e. for a fixed a, the index k' runs only over $1, \cdots, e^{S_0(M-E_a)} \ll e^{S_0(M)}$. The combination of Eqs. (2.34, 2.35) then yields

$$|\psi_k(M)\rangle|d_0\rangle \to \sum_{i,k_i} \alpha^k_{ik_i} |\psi_{k_i}(M-E_i)\rangle|d_i\rangle, \tag{2.36}$$

where $\alpha^k_{ik_i} = \sum_{a,k'} c^k_{iak'} d^{ak'}_{k_i}$ and $E_i = E_{\mathrm{d},i}\sqrt{1-2Ml_{\mathrm{P}}^2/r_{\mathrm{d}}}$. This represents a microscopic, unitary description of the elementary mining process.

In the description given above, we have separated the detector state from the state of the black hole, but in a treatment fully consistent with the notation in earlier sections, the detector itself must be viewed as excitations over $|\psi_k(M)\rangle$. After the detector response process in Eq. (2.34), these excitations can be entangled with Hawking quanta emitted earlier, reflecting the fact that the detector can extract information from the black hole. Since the detector can now be put deep in the zone, in which the Rindler approximation is applicable, this implies that excitations localized within the Rindler wedge corresponding to the region $r > r_{\mathrm{s}}$ are entangled with early Hawking radiation. Does this lead to firewalls as discussed in Ref. [10]? The answer is no. The excitations describing the detector are, in the near horizon Rindler approximation, those of modes that are smooth in Minkowski space (a modes in the notation of Ref. [10]). Likewise, modes representing negative energy excitations arising from the backreactions are also ones smooth in Minkowski space. Excitations of these modes, of course, *do* perturb the black hole system, which can indeed be significant if the detector is held very close to the horizon. This effect, however, is caused by physical interactions between the detector and vacuum degrees of freedom, and is confined in the causal future of the interaction event. This is not the firewall phenomenon.

The semiclassical description of the mining process in Eq. (2.36) is obtained by taking maximal mixture for the vacuum indices. Specifically, the semiclassical state before the process starts is given by

$$\rho(M) = \frac{1}{e^{S_0(M)}} \sum_{k=1}^{e^{S_0(M)}} |\psi_k(M)\rangle|d_0\rangle\langle\psi_k(M)|\langle d_0|. \tag{2.37}$$

The evolution of this state under Eq. (2.36) is then

$$\rho(M) \to \frac{1}{e^{S_0(M)}} \sum_{k=1}^{e^{S_0(M)}} \sum_{i,i'} \sum_{k_i=1}^{e^{S_0(M-E_i)}} \sum_{k'_{i'}=1}^{e^{S_0(M-E_{i'})}} \alpha^k_{ik_i} \alpha^{k*}_{i'k'_{i'}} |\psi_{k_i}(M-E_i)\rangle |d_i\rangle \langle \psi_{k'_{i'}}(M-E_{i'})| \langle d_{i'}|. \tag{2.38}$$

This leads to the density matrix describing the detector state after the process

$$\rho_{\rm d} = \sum_{i,i'} \gamma_{ii'} \, |d_i\rangle \langle d_{i'}|, \tag{2.39}$$

where

$$\gamma_{ii'} = {\rm Tr}\left[\frac{1}{e^{S_0(M)}} \sum_{k=1}^{e^{S_0(M)}} \sum_{k_i=1}^{e^{S_0(M-E_i)}} \sum_{k'_{i'}=1}^{e^{S_0(M-E_{i'})}} \alpha^k_{ik_i} \alpha^{k*}_{i'k'_{i'}} |\psi_{k_i}(M-E_i)\rangle \langle \psi_{k'_{i'}}(M-E_{i'})| \right]. \tag{2.40}$$

Assuming that the microscopic dynamics of the vacuum degrees of freedom are generic, $\gamma_{ii'}$ is expected to take the form

$$\gamma_{ii'} \approx \frac{1}{Z} f_i e^{-\frac{E_{{\rm d},i}}{T(r_{\rm d})}} \, \delta_{ii'}, \tag{2.41}$$

where $Z = \sum_i f_i e^{-E_{{\rm d},i}/T(r_{\rm d})}$, and f_i is the detector response function reflecting intrinsic properties of the detector under consideration. This implies that in the semiclassical approximation, the final detector state does not have any information about the original black hole microstate, despite the fact that the fundamental process in Eq. (2.36) is, in fact, unitary.

The fate of an infalling object

We now discuss how an object falling into a black hole is described in a distant reference frame. As we have seen, having a well-defined black hole geometry requires a superposition of an enormous number of energy-momentum eigenstates. While the necessary spreads in energy and momentum are small when measured in the asymptotic region, the spreads of *local* energy and momentum (i.e. those measured by local approximately static observers) are large in the region close to the horizon, because of large gravitational blueshifts. This makes the local temperature $T(r)$ associated with the vacuum degrees of freedom, Eq. (3.7), very high near the horizon. We expect that the semiclassical description becomes invalid when this temperature exceeds the string (cutoff) scale, $T(r) \gtrsim 1/l_*$. Namely, semiclassical spacetime exists only in the region

$$r > r_{\rm s} = 2Ml_{\rm P}^2 + O\left(\frac{l_*^2}{Ml_{\rm P}^2}\right), \tag{2.42}$$

where $r_{\rm s}$ is identified as the location of the stretched horizon. The same conclusion can also be obtained by demanding that the gravitational thermal entropy stored in the region where

the semiclassical spacetime picture is applicable is a half of the Bekenstein-Hawking entropy, $\mathcal{A}/8l_{\rm P}^2$, as discussed in footnote 13.

Let us consider that an object is dropped from $r = r_0$ with vanishing initial velocity, where $r_0 - 2Ml_{\rm P}^2 \approx O(Ml_{\rm P}^2) > 0$. It then freely falls toward the black hole and hits the stretched horizon at $r = r_{\rm s}$ in Schwarzschild time of about $4Ml_{\rm P}^2 \ln(Ml_{\rm P}^2/l_*)$. Before it hits the stretched horizon, the object is described by a and $a_{\rm far}$, the indices labeling field and string theoretic excitations over the semiclassical background spacetime. After hitting the stretched horizon, the information about the object will move to the index $\bar{a}$, labeling excitations of the stretched horizon. The information about the fallen object will then stay there, at least, for the thermalization (or scrambling) time of the stretched horizon, of order $Ml_{\rm P}^2 \ln(Ml_{\rm P})$. This allows for avoiding the inconsistency of quantum cloning in black hole physics [88, 161]. Finally, the information in $\bar{a}$ will further move to k, which can (later) be extracted by an observer in the asymptotic region via the Hawking emission or mining process, as described in the previous two subsections.

We note that the statement that an object is in the semiclassical regime (i.e. represented by indices a and $a_{\rm far}$) does *not* necessarily mean that it is well described by semiclassical *field* theory. Specifically, it is possible that stringy effects become important before the object hits the stretched horizon. As an example, consider dropping an elementary particle of mass m ($\ll 1/l_*$) from $r = r_0$ with zero initial velocity. (Here, by elementary we mean that there is no composite structure at lengthscale larger than l_*.) The local energy and local radial momentum of the object will then vary, as it falls, as:

$$E_{\rm loc} = m\sqrt{\frac{1-\frac{2Ml_{\rm P}^2}{r_0}}{1-\frac{2Ml_{\rm P}^2}{r}}}, \qquad p_{\rm loc} = -m\sqrt{\frac{\frac{2Ml_{\rm P}^2}{r}-\frac{2Ml_{\rm P}^2}{r_0}}{1-\frac{2Ml_{\rm P}^2}{r}}}. \tag{2.43}$$

The values of $E_{\rm loc} \approx -p_{\rm loc}$ get larger as r gets smaller, and for $m \gg 1/Ml_{\rm P}^2$ (which we assume here) become of order $1/l_*$ before the object hits the stretched horizon, i.e. at

$$r - 2Ml_{\rm P}^2 \simeq 2Ml_{\rm P}^2(ml_*)^2\left(1-\frac{2Ml_{\rm P}^2}{r_0}\right). \tag{2.44}$$

The Schwarzschild time it takes for the object to reach this point is only about $-4Ml_{\rm P}^2 \ln(ml_*)$, much smaller than the time needed to reach the stretched horizon, $4Ml_{\rm P}^2 \ln(Ml_{\rm P}^2/l_*)$. After the object reaches this point, i.e. when $E_{\rm loc} \approx -p_{\rm loc} \gtrsim 1/l_*$, stringy effects might become important; specifically, its Lorentz contraction saturates and transverse size grows with $E_{\rm loc}$ [167]. Note that this dependence of the description on the boost of a particle does not necessarily mean violation of Lorentz invariance—physics can still be fully Lorentz invariant.[18]

[18] It is illuminating to consider how these stringy effects appear in a two-particle scattering process in Minkowski space. For $\sqrt{s} \lesssim 1/l_*$, where s is the Mandelstam variable, there is a reference frame in which energies/momenta of *both* particles are smaller than $1/l_*$, guaranteeing that these effects are not important in the process. For $\sqrt{s} > 1/l_*$, on the other hand, at least one particle has an energy/momentum larger than $1/l_*$ in *any* reference frame, suggesting that stringy effects become important in scattering with such high $\sqrt{s}$.

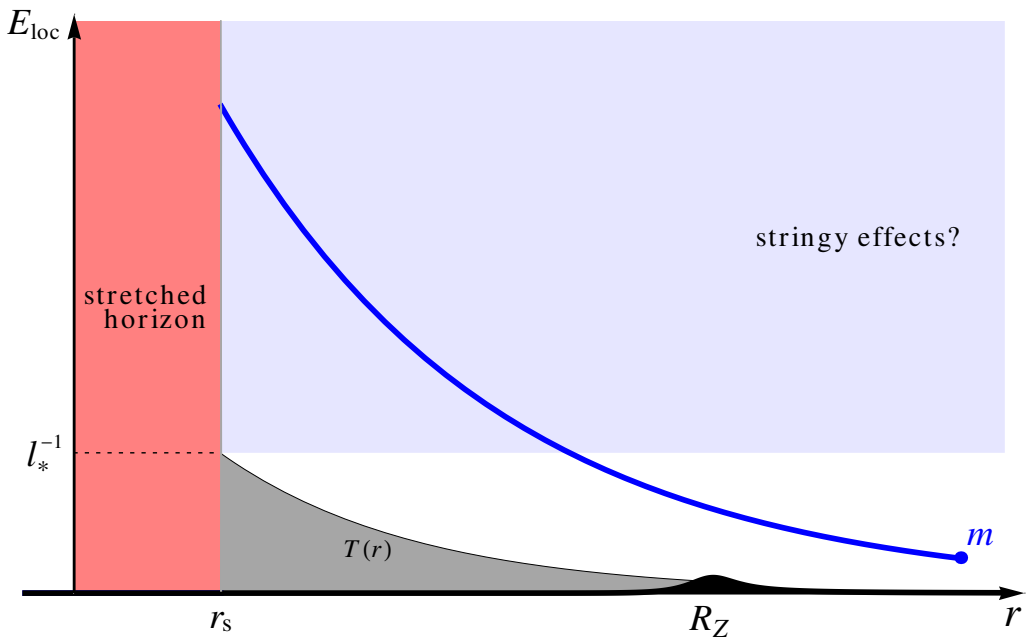


Figure 2.3: A schematic depiction of the fate of an elementary particle of mass m ($1/Ml_{\rm P}^2 \ll m \ll 1/l_*$) dropped into a black hole, viewed in a distant reference frame. As the particle falls, its local energy blueshifts and exceeds the string/cutoff scale $1/l_*$ before it hits the stretched horizon. After this point, stringy effects could become important, although the semiclassical description of the object may still be applicable. The object hits the stretched horizon at a Schwarzschild time of about $4Ml_{\rm P}^2 \ln(Ml_{\rm P}^2/l_*)$ after the drop. After this time, the semiclassical description of the object is no longer applicable, and the information about the object will be encoded in the index $\bar{a}$, representing excitations of the stretched horizon. (This information will further move to the vacuum index k later, so that it can be extracted by an observer in the asymptotic region via the Hawking emission or mining process.)

A schematic picture for the fate of an infalling object described above is given in Fig. 2.3. In a distant reference frame, the semiclassical description of the object is applicable only until it hits the stretched horizon, after which it is represented as excitations of the stretched horizon. On the other hand, according to general relativity (or the equivalence principle), the falling object does not experience anything other than smooth empty spacetime when it crosses the horizon, except for effects associated with curvature, which are very small for a black hole of mass $M \gg 1/l_{\rm P}$. If this picture is correct, then we expect there is a way to reorganize the dynamics of the stretched horizon such that the general relativistic smooth interior of the black hole becomes manifest. In the complementarity picture, this is achieved by performing an appropriate reference frame change. We now move on to discuss this issue.

2.4 Black Hole—An Infalling Description

In order to describe the fate of an infalling object using low energy language after it crosses the Schwarzschild horizon, we need to perform a change of the reference frame from a distant one, which we have been considering so far, to an infalling one which falls into the black hole with the object. In general, studying this issue is complicated by the fact that the general and precise formulation of complementarity is not yet known, but we may still explore the expected physical picture based on some general considerations.

The aim of this section is to argue that the existence of interior spacetime, as suggested by general relativity, does not contradict the unitarity of the Hawking emission and black hole mining processes, as described in the previous section in a distant reference frame. We do this by first arguing that there exists a reference frame—an infalling reference frame—in which the spacetime around a point on the Schwarzschild horizon appears as a large nearly flat region, with the curvature lengthscale of order $Ml_{\rm P}^2$. This is a reference frame whose origin falls freely from rest from a point sufficiently far from the black hole. We discuss how the description based on this reference frame is consistent with that in the distant reference frame, despite the fact that they apparently look very different, for example in spacetime locations of the vacuum degrees of freedom.

We then discuss how the system is described in more general reference frames, in particular a reference frame whose origin falls from rest from a point close to the Schwarzschild horizon. We will also discuss (non-)relations of black hole mining by a near-horizon static detector and the—seemingly similar—Unruh effect in Minkowski space. The discussion in this section illuminates how general coordinate transformations may work at the level of full quantum gravity, beyond the approximation of quantum field theory in curved spacetime.

Emergence of interior spacetime—free fall from a distance

What does a reference frame really mean? According to the general complementarity picture described in Section 2.2, it corresponds to a foliation of a portion of spacetime which a single (hypothetical) observer can access. As discussed there, the procedure to erect such a reference frame should not depend on the background geometry in order for the framework to be applicable generally, and there is currently no precise, established formulation to do that (although there are some partially successful attempts; see, e.g., Ref. [140]). Here we focus only on classes of reference frames describing the same system with a fixed black hole background. This limitation allows us to bypass many of the issues arising when we consider the most general application of the complementarity picture.

In this subsection, we consider a class of reference frames which we call infalling reference frames. We argue that a reference frame in this class makes it manifest that the spacetime near the origin of the reference frame appears as a large approximately flat region when it crosses the Schwarzschild horizon, up to corrections from curvature of lengthscale $Ml_{\rm P}^2$. We discuss how the interior spacetime of the black hole can emerge through the complementarity transformation representing a change of reference frame from the distant to infalling ones.

Consistency of the infalling picture described here with the distant frame description in Section 2.3 will be discussed in more detail in the next subsection.

We consider a reference frame associated with a freely falling (local Lorentz) frame, with its spatial origin p_0 following the worldline representing a hypothetical observer [133, 140]. In particular, we let the origin of the reference frame, p_0, follow the trajectory of a timelike geodesic, representing the observer who is released from rest at $r = r_0$, with r_0 sufficiently far from the Schwarzschild horizon, $r_0 - 2Ml_{\rm P}^2 \gtrsim Ml_{\rm P}^2$. According to the complementarity hypothesis, the system described in this reference frame does not have a (hot) stretched horizon at the location of the Schwarzschild horizon when p_0 crosses it. (The stretched horizon must have existed around the Schwarzschild horizon when p_0 was far away, $r_{p_0} - 2Ml_{\rm P}^2 \gtrsim O(Ml_{\rm P}^2)$, because the description in those earlier times must be approximately that of a distant reference frame, i.e. that discussed in the previous section.) In particular, the region around p_0 must appear approximately flat, i.e. up to small effects from curvature of order $1/M^2 l_{\rm P}^4$, until p_0 approaches the singularity.

In this infalling description, we expect that a "horizon" signaling the breakdown of the semiclassical description lies in the directions associated with "past-directed and inward" light rays (the directions with increasing r and decreasing t after p_0 crosses $r = 2Ml_{\rm P}^2$) as viewed from p_0; see Fig. 2.4.[19] As in the stretched horizon in a distant reference frame, this "horizon" emerges because of the "squeezing" of equal-time hypersurfaces; in particular, an observer following the trajectory of p_0 may probe only a tiny region near the Schwarzschild horizon for signals arising from this surface. (Note that $-r$ plays a role of time inside the Schwarzschild horizon.) Considering angular directions, this "horizon" has an area of order $M^2 l_{\rm P}^4$, and can be regarded as being located at distances of order $Ml_{\rm P}^2$ away from p_0 (with an appropriately defined distance measure on generic equal-time hypersurfaces in the infalling reference frame; see Section 2.4).

In analogy with the case of a distant frame description, we denote basis states for the general microstates in an infalling reference frame (before p_0 reaches the singularity) as

$$|\Psi_{\bar{\alpha}\,\alpha\,\alpha_{\rm far};\kappa}(M)\rangle, \tag{2.45}$$

where $\bar{\alpha}$ labels the excitations of the "horizon," and α, and $\alpha_{\rm far}$ are the indices labeling the semiclassical excitations near and far from the black hole, conveniently defined; κ is the vacuum index in an infalling reference frame, representing degrees of freedom that cannot be resolved by semiclassical operators.[20] The complementarity transformation provides a map from the basis states in a distant description, Eq. (3.3), to those in an infalling description, Eq. (3.13), and vice versa. The general form of this transformation can be quite complicated, depending, e.g., on equal-time hypersurfaces taken in the two descriptions (which are in

[19] This "horizon," as viewed from an infalling reference frame, should not be confused with the stretched, or Schwarzschild, horizon as viewed from a distant reference frame.

[20] After p_0 hits the singularity, the system as viewed from the infalling reference frame can only be represented by "singularity states": intrinsically quantum gravitational states that do not allow for a spacetime interpretation [133].

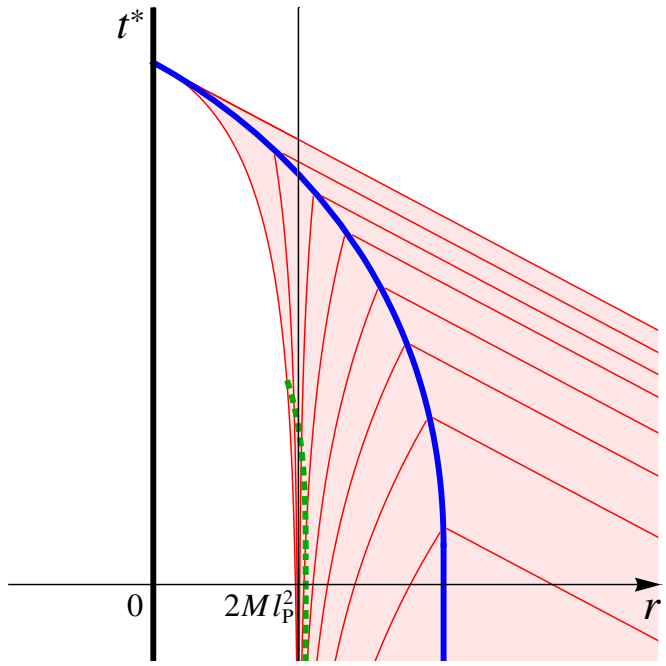


Figure 2.4: A sketch of an infalling reference frame in an Eddington-Finkelstein diagram: the horizontal and vertical axes are r and $t^* = t + r^* - r$, respectively, where r^* is the tortoise coordinate. The thick (blue) line denotes the spacetime trajectory of the origin, p_0, of the reference frame, while the thin (red) lines represent past-directed light rays emitted from p_0. The shaded area is the causal patch associated with the reference frame, and the dotted (green) line represents the stretched "horizon" as viewed from this reference frame.

turn related with the general procedure of erecting reference frames by standard coordinate transformations within each causal patch). Here we consider how various indices are related under the transformation, focusing on the near black hole region.

Imagine that equal-time hypersurfaces in the two—distant and infalling—reference frames agree at some time $t = t_0$ in the spacetime region near but outside the surface where the stretched horizon exists if viewed from the distant reference frame. (Note that the stretched horizon has physical substance only in a distant reference frame.) We are interested in how basis states in the two descriptions transform between each other in the timescale of the fall of the infalling reference frame. The time here can be taken as the proper time at p_0 in each reference frame [133, 140], which is approximately the Schwarzschild time for the distant reference frame. In this case, the relevant timescale is $t - t_0 \lesssim O(Ml_P^2 \ln(Ml_P))$ in the distant reference frame, while $t - t_0 \lesssim O(Ml_P^2)$ in the infalling reference frame.

As discussed in Section 2.3, in the distant reference frame, an object dropped from some r_0 with $r_0 - 2Ml_P^2 \approx O(Ml_P^2)$ is first represented by a and then by $\bar{a}$ after it hits the stretched horizon. On the other hand, in the infalling frame, the object is represented by the index α throughout, first as a semiclassical excitation outside the Schwarzschild horizon and then as

a semiclassical excitation inside the Schwarzschild horizon, implying that the object does not find anything special at the horizon. Here, we have assumed that p_0 follows (approximately) the trajectory of the falling object. This suggests that a portion of the α index representing excitations in the interior of the black hole is transformed into the $\bar{a}$ index in the distant description (and vice versa) under the complementarity transformation; i.e., the interior of the black hole accessible from the infalling reference frame is encoded in the excitations of the stretched horizon in the distant reference frame. Note that the amount of information needed to reconstruct the interior (in the semiclassical sense) is much smaller than the Bekenstein-Hawking entropy [173, 142]—the logarithm of the dimension of the relevant Hilbert space is of order $(\mathcal{A}/l_{\rm P}^2)^q$ with $q < 1$.

In the exterior spacetime region, the portion of the α index representing excitations there, as well as the $\alpha_{\rm far}$ index, are mapped to the corresponding a and $a_{\rm far}$ indices, and vice versa (after matching the equal-time hypersurface in the two descriptions through appropriate time evolutions). Because equal-time hypersurfaces foliate the causal patch, excitations in the far exterior region naturally have trans-Planckian energies in the infalling description. However, as discussed in Section 2.3, this does not mean that the semiclassical description is invalid—objects may still be described as excitations in the semiclassical spacetime, although stringy effects may become important. Indeed, we expect that the semiclassical description is applicable in the far exterior region even in the infalling reference frame, because of the absence of the "squeezing" effect described above which leads to the breakdown of the semiclassical picture.

We emphasize that the construction of the interior spacetime described here does not suffer from the paradoxes discussed in Refs. [10, 9, 126]. By labeling states in terms of excitations, we are in a sense representing the interior spacetime already in the distant description. (The interpretation, however, is different. In the distant description, the relevant excitations must be regarded as those of the stretched horizon.) In fact, we do not find any inconsistency in postulating that the dynamics of an infalling object is described by the corresponding Hamiltonian in the semiclassical theory in a sufficiently small region around p_0, to the extent that microscopic details of interactions with κ degrees of freedom are neglected. Namely, we do not find any inconsistency in postulating that physics at the classical level is well described by general relativity.

Finally, we discuss where the fine-grained vacuum degrees of freedom represented by κ must be viewed as being located in the infalling description. Because of the lack of an obvious static limit, it is not straightforward to answer to this question. Nevertheless, it seems natural to expect, in analogy with the case of a distant description, that most of the degrees of freedom are located close to the "horizon" (in terms of a natural distance measure in which the distance between the "horizon" and p_0 is of order $Ml_{\rm P}^2$). In fact, we expect that the number of κ degrees of freedom existing around p_0 within a distance scale sufficiently smaller than $Ml_{\rm P}^2$ is of $O(1)$ or smaller, since the time and length scales of the system characterizing local deviations from Minkowski space (as viewed from the infalling reference frame) are both of order $Ml_{\rm P}^2$. As in the case of the distant description, we expect that the κ degrees of freedom do not extend significantly to the far exterior region, since the

existence of the black hole does not affect the spacetime there much.[21]

Consistency between the distant and infalling descriptions

In analyzing a black hole system in a distant reference frame, we argued that the microscopic information about the black hole, represented by the k index, is distributed according to the gravitational thermal entropy calculated using semiclassical field theory. In particular, on the Schwarzschild (or stretched) horizon, this information has a Planckian density: one qubit per area of order $l_{\rm P}^2$ on the horizon (or per volume of order $l_{\rm P}^3$ if we take into account the "thickness" of the stretched horizon, $\sim l_{\rm P}$). On the other hand, we have just argued that in an infalling reference frame, the spacetime distribution of the microscopic information (now represented by the κ index) is different. In particular, the spatial density of the information around the Schwarzschild horizon, when the origin of the reference frame passes through it, is very small: one qubit per volume of order $(Ml_{\rm P}^2)^3$. How can we reconcile these two seemingly very different perspectives?

In this subsection, we consider this problem and argue that despite the fact that the spacetime distribution of the microscopic information depends on the reference frame one chooses to describe the system, the answers to any operationally well-defined question one obtains in different reference frames are consistent with each other. As an example most relevant to our discussion, we consider a physical detector hovering at a constant Schwarzschild radius $r = r_{\rm d}$ ($> 2Ml_{\rm P}^2$). In a distant description, the spatial density of the microscopic information, represented by k, is large at the location of the detector when $r_{\rm d} - 2Ml_{\rm P}^2 \ll Ml_{\rm P}^2$. Such a detector (or a system of detectors) can thus be used for black hole mining: accelerated extraction of energy and information from the black hole. In an infalling reference frame, however, the density of the microscopic information, represented by κ, is very small at the detector location, at least when the origin of the reference frame, p_0, passes nearby. This implies that the rate of extracting information from spacetime cannot be much faster than $1/Ml_{\rm P}^2$ around p_0 in the infalling description, reflecting the fact that the spacetime appears approximately flat there. How are these two descriptions consistent?

In the distant description, the rate of extracting microscopic information about the black hole is at most of order one qubit per Schwarzschild time $1/T_{\rm H} = 8\pi Ml_{\rm P}^2$ *per channel*, regardless of the location of the detector [39]—the acceleration of information extraction occurs not because of a higher speed of information extraction in each channel but because of an increased number of channels available by immersing the detector deep into the zone. This implies that each single detector, which we define to act on a single channel, "clicks" once (i.e. extracts of $O(1)$ qubits) per a Schwarzschild time of order $8\pi Ml_{\rm P}^2$.

[21] Note that the descriptions in the two reference frames are already different at the semiclassical level. For example, the backreaction of a detector click in a distant reference frame is described as an absorption of a particle in the thermal bath, while in an infalling reference frame it is described as an emission of a particle, with the difference arising from different definitions of energy in the two reference frames [181]. The reference frame dependence discussed here is much more drastic, however—the spacetime locations of physical degrees of freedom are different in the two reference frames.

Now, consider describing such a detector in an infalling reference frame whose origin p_0 is released at $r = 2Ml_{\rm P}^2 + O(Ml_{\rm P}^2)$ from rest, at an angular location close to the detector. To understand the relevant kinematics, we adopt the near-horizon Rindler approximation: for $r > 2Ml_{\rm P}^2$

$$\rho \approx 2\sqrt{2Ml_{\rm P}^2(r - 2Ml_{\rm P}^2)}, \qquad \omega \approx \frac{t}{4Ml_{\rm P}^2}, \tag{2.46}$$

in terms of which the metric is given by

$$ds^2 \approx -\rho^2 d\omega^2 + d\rho^2 + r(\rho)^2 d\Omega. \tag{2.47}$$

As is well-known, this metric can be written in the Minkowski form

$$ds^2 \approx -dT^2 + dZ^2 + r(T,Z)^2 d\Omega, \tag{2.48}$$

by introducing the coordinates

$$T = \rho \sinh \omega, \qquad Z = \rho \cosh \omega, \tag{2.49}$$

which can be extended into the $r < 2Ml_{\rm P}^2$ region. Our setup corresponds to the situation in which the detector follows a trajectory of a constant ρ:

$$\rho = \rho_{\rm d} \ll Ml_{\rm P}^2, \tag{2.50}$$

while the origin of the reference frame p_0—or the (fictitious) observer—is at a constant Z:

$$Z = Z_{\rm o} \approx O(Ml_{\rm P}^2). \tag{2.51}$$

Note that while we *approximate* the geometry by flat space, given by Eq. (2.47) or (2.48), the actual system has small nonzero curvature with lengthscale of order $Ml_{\rm P}^2$.

As discussed above, the detector extracts an $O(1)$ amount of information in each time interval of

$$\Delta\omega \approx O\left(\frac{1}{4Ml_{\rm P}^2 T_H}\right) \approx O(1), \tag{2.52}$$

while the "observer," p_0, and the detector meet (or pass by each other) at

$$\begin{pmatrix} \omega \\ \rho \end{pmatrix} = \begin{pmatrix} \operatorname{arccosh}\frac{Z_{\rm o}}{\rho_{\rm d}} \\ \rho_{\rm d} \end{pmatrix} \equiv \begin{pmatrix} \omega_* \\ \rho_* \end{pmatrix}. \tag{2.53}$$

This implies that in the Minkowski coordinates—i.e. as viewed from the infalling observer p_0—the detector clicks only once in each time/space interval of

$$\Delta T \approx \Delta\omega \left.\frac{\partial T}{\partial \omega}\right|_{\omega=\omega_*,\rho=\rho_*} \approx Z_{\rm o} \approx O(Ml_{\rm P}^2), \tag{2.54}$$

$$\Delta Z \approx \Delta\omega \left.\frac{\partial Z}{\partial \omega}\right|_{\omega=\omega_*,\rho=\rho_*} \approx Z_{\rm o} \approx O(Ml_{\rm P}^2), \tag{2.55}$$

around p_0. This is precisely what we expect from the equivalence principle: the spacetime appears approximately flat when viewed from an infalling observer, up to curvature effects with lengthscale of $Ml_{\rm P}^2$. While the detector clicks of order $\ln(Ml_{\rm P})$ times within the causal patch of the infalling reference frame, all these clicks occur at distances of order $Ml_{\rm P}^2$ away from p_0, where we expect a higher density of κ degrees of freedom. The two descriptions—distant and infalling—are therefore consistent, despite the fact that the spacetime distributions of the microscopic information about the black hole—represented by k and κ, respectively—are different in the two reference frames.

While we have so far discussed the case in which a physical detector is located close to the Schwarzschild horizon, the conclusion is the same in the case of spontaneous Hawking emission. In this case, since Hawking particles appear as semiclassical excitations only at $r - 2Ml_{\rm P}^2 \gtrsim Ml_{\rm P}^2$ with local energies of order $1/Ml_{\rm P}^2$, the consistency of the two descriptions is in a sense obvious. Alternatively, one can regard this case as the $\rho_{\rm d} \approx Ml_{\rm P}^2$ limit of the previous analysis. While the Rindler approximation is strictly valid only for ρ sufficiently smaller than $Ml_{\rm P}^2$, qualitative results are still valid for $\rho_{\rm d} \approx Ml_{\rm P}^2$; in particular, the estimates in Eqs. (2.54, 2.55) are valid at an order of magnitude level.

Other reference frames—free fall from a nearby point

In this subsection, we consider how the black hole is described in a class of reference frames whose origin follows a timelike geodesic released from rest at $r = r_0$, where r_0 is *close to* the Schwarzschild horizon, $r_0 - 2Ml_{\rm P}^2 \ll Ml_{\rm P}^2$.[22] We argue that the description in these reference frames does not look similar to either the distant or infalling description discussed before, and yet it is consistent with both of them.[23]

To understand how the black hole appears in such a reference frame, let us consider a setup similar to that in Section 2.4—a physical detector hovering at a constant Schwarzschild radius $r = r_{\rm d}$—and see how this system is described in the reference frame. As in Section 2.4, we may adopt the Rindler approximation, in which Eq. (2.51) is now replaced by

$$Z = Z_{\rm o} \ll Ml_{\rm P}^2. \tag{2.56}$$

This implies that as viewed from this reference frame, the detector clicks once in each time/space interval of

$$\Delta T \approx \Delta Z \approx Z_{\rm o} \ll Ml_{\rm P}^2. \tag{2.57}$$

Here, we have assumed that $\rho_{\rm d} < Z_{\rm o}$. Since each detector click extracts an $O(1)$ amount of information from spacetime, which we expect not to occur in Minkowski space, this implies

[22] In a full geometry in which the black hole is formed by collapsing matter, the trajectory of the origin, p_0, of such a reference frame corresponds to a fine-tuned one in which p_0 stays near outside of the Schwarzschild horizon for long time due to large outward velocities at early times. (Here, we have focused only on the relevant branch in the full quantum state; see, e.g., footnote 4.)

[23] Note that we use the term "infalling reference frame" exclusively for reference frames discussed in Sections 2.4 and 2.4, i.e. the ones in which p_0 starts from rest at r_0 with $r_0 - 2Ml_{\rm P}^2 \gtrsim O(Ml_{\rm P}^2)$.

that the spacetime cannot be viewed as approximately Minkowski space over a region beyond lengthscale $Z_{\rm o}$. In particular, in contrast with the case in an infalling reference frame (with $Z_{\rm o} \gtrsim O(Ml_{\rm P}^2)$), the spacetime region around p_0 in this reference frame does not appear nearly flat over lengthscale of $Ml_{\rm P}^2$ when p_0 crosses the Schwarzschild horizon.

At a technical level, this difference arises from the fact that the relative boost of p_0 with respect to the distant reference frame when p_0 approaches the detector

$$\gamma = \frac{1}{\sqrt{1-v_{\rm rel}^2}} = \sqrt{\frac{1-\frac{2Ml_{\rm P}^2}{r_0}}{1-\frac{2Ml_{\rm P}^2}{r_d}}}, \tag{2.58}$$

is very different in the two reference frames. In an infalling reference frame γ is huge, $\approx O(Ml_{\rm P}^2/\rho_{\rm d})$, while in the reference frame considered here $\gamma \approx O(Z_{\rm o}/\rho_{\rm d})$, which is not as large as that in the infalling case. In the infalling reference frame of Sections 2.4 and 2.4, the huge boost of $\gamma \approx O(Ml_{\rm P}^2/\rho_{\rm d})$ "stretched" the interval between detector clicks to time/length scales of order $Ml_{\rm P}^2$. Here, this "stretching" makes only a small region around p_0, with lengthscale of order $Z_{\rm o}$ ($\ll Ml_{\rm P}^2$), look nearly flat at any given time.

We may interpret this result to mean that in the reference frame under consideration, the "horizon" (as viewed from this reference frame) is located at a distance of order $Z_{\rm o}$ away from p_0, so that detector clicks occur near or "on" this surface. (In the latter case, the detector click events must be viewed as occurring in the regime outside the applicability of the semiclassical description; in particular, they can only be described as complicated quantum gravitational processes occurring on the "horizon.") Since we expect that microscopic information about the black hole (analogous to k and κ in the distant and infalling reference frames, respectively) is located near and on the "horizon," there is no inconsistency that detector clicks extract microscopic information from the black hole.

One might be bothered by the fact that in this reference frame spacetime near the Schwarzschild horizon does not appear large, $\approx O(Ml_{\rm P}^2)$, nearly flat space, and consider that this implies the non-existence of a large black hole interior as suggested by general relativity. This is, however, not correct. The existence of *a* reference frame in which spacetime around the Schwarzschild horizon appears as a large nearly flat region—in particular, the existence of an infalling reference frame discussed in Sections 2.4 and 2.4—already ensures that an infalling physical object/observer does not experience anything special, e.g. firewalls, when it/he/she crosses the Schwarzschild horizon. The analysis given here simply says that the spacetime around the Schwarzschild horizon does *not always* appear as a large nearly flat region, even in a reference frame whose origin falls freely into the black hole. This extreme relativeness of descriptions is what we expect from complementarity.

(Non-)relations with the Unruh effect in Minkowski space

It is often thought that the system described above is similar to an accelerating detector existing in Minkowski space, based on a similarity of geometries between the two setups. If this were true at the full quantum level, it would mean that the description in an *inertial*

reference frame in Minkowski space must possess a "horizon," at which the semiclassical description of the system breaks down. Does this make sense?

Here we argue that physics of a detector held near the Schwarzschild horizon, given above in Section 2.4, is, in fact, different from that of an accelerating detector in Minkowski space. The intuition that the two must be similar comes from the (wrong) perception that the detector located near the Schwarzschild horizon feels a high blueshifted Hawking temperature, $\approx 1/\rho_{\rm d} \gg 1/Ml_{\rm P}^2$, which makes the detector click at a high rate, while the spacetime curvature there is very small, with lengthscale $\approx Ml_{\rm P}^2$, so that such a tiny curvature must not affect the system. This intuition, however, is flawed by mixing up two different pictures—the system as viewed at the location of the detector and as viewed in the asymptotic region.

Suppose we represent all quantities as defined in the asymptotic region. The temperature a detector feels is then of order $1/Ml_{\rm P}^2$ and the timescale for detector clicks is $T \approx O(Ml_{\rm P}^2)$ for *any* $r_{\rm d} > 2Ml_{\rm P}^2$. On the other hand, the energy density of the black hole region is of order $M/(Ml_{\rm P}^2)^3$, so that the curvature lengthscale L is estimated as

$$\frac{1}{L^2} \sim G_{\rm N}\frac{M}{(Ml_{\rm P}^2)^3} \sim \frac{1}{(Ml_{\rm P}^2)^2}. \tag{2.59}$$

This implies that

$$T \sim L \sim O(Ml_{\rm P}^2); \tag{2.60}$$

namely, curvature is expected to give *an* $O(1)$ *effect* on the dynamics of the detector response.

The same conclusion can also be reached when we represent all the quantities in the static frame at the detector location. In this case, the temperature the detector feels is of order $1/Ml_{\rm P}^2\chi$, where $\chi = \sqrt{1-2Ml_{\rm P}^2/r_{\rm d}}$ is the redshift factor, so that $T \approx O(Ml_{\rm P}^2\chi)$. On the other hand, the energy density of the black hole region is given by $\sim (M/\chi)/(Ml_{\rm P}^2)^3\chi$, so that the "blueshifted curvature length" L is given by

$$\frac{1}{L^2} \sim G_{\rm N}\frac{M/\chi}{(Ml_{\rm P}^2)^3\chi} \sim \frac{1}{(Ml_{\rm P}^2\chi)^2}. \tag{2.61}$$

This yields

$$T \sim L \sim O(Ml_{\rm P}^2\chi), \tag{2.62}$$

again implying that curvature provides an $O(1)$ effect on the dynamics.

It is, therefore, no surprise that the physics of a near-horizon detector in Section 2.4 differs significantly from that of an accelerating detector in Minkowski space experiencing the Unruh effect [179]. In fact, we consider, as we naturally expect, that an inertial frame description in Minkowski space does *not* have a horizon, implying that no information about spacetime is extracted by an accelerating detector, despite the fact that it clicks at a rate controlled by the acceleration a, $T \approx O(1/a)$, in the detector's own frame. This is indeed consistent with the idea that any information must be accompanied by energy. In the black hole case, the detector mines the black hole, i.e. its click extracts energy from the black

hole spacetime, while in the Minkowski case the energy needed to excite the detector comes entirely from the force responsible for the acceleration of the detector—the detector does not mine energy from Minkowski space. We conclude that blueshifted Hawking radiation and Unruh radiation in Minkowski space are very different as far as the information flow is concerned.

Does this imply a violation of the equivalence principle? The equivalence principle states that gravity is the same as acceleration, and the above statement might seem to contradict this principle. This is, however, not true. The principle demands the equivalence of the two only at a point in space in a given coordinate system, and the descriptions of the two systems discussed above—a black hole and Minkowski space—are indeed the same in an infinitesimally small (or lengthscale of order l_*) neighborhood of p_0. The principle does not require that the descriptions must be similar in regions away from p_0, and indeed they are very different: there is a "horizon" at a distance of order $Z_{\rm o}$ from p_0 in the black hole case while there is no such thing in the Minkowski case. And it is precisely in these regions that the detector clicks to extract (or non-extract) information from the black hole (Minkowski) spacetime. In quantum mechanics, a system is specified by a quantum state which generally encodes global information on the equal-time hypersurface. It is, therefore, natural that the equivalence principle, which makes a statement only about a point, does not enforce the equivalence between physics of blueshifted Hawking radiation and of the Unruh effect in Minkowski space at the fully quantum level.

Complementarity: general covariance in quantum gravity

We have argued that unitary information transfer described in Section 2.3, associated with Hawking emission and black hole mining, is consistent with the existence of the interior spacetime suggested by general relativity. We can summarize important lessons we have learned about quantum gravity through this study in the following three points:

- In *a fixed reference frame*, the microscopic information about spacetime, in this case about a black hole, may be viewed as being associated with specific spacetime locations. In particular, for a (quasi-)static description of a system, these degrees of freedom are distributed according to the gravitational thermal entropy calculated using semiclassical field theory. The distribution of these degrees of freedom—which we may call "constituents of spacetime"—controls how they can interact with the degrees of freedom in semiclassical theory, e.g. matter and radiation in semiclassical field theory.

- The spacetime distribution of the microscopic information, however, changes if we adopt a different reference frame to describe the system. In this sense, the "constituents of spacetime" are *not* anchored to spacetime; they are associated with specific spacetime locations only after the reference frame is fixed. In particular, no reference frame independent statement can be made about where these degrees of freedom are located in spacetime. We may view this as a manifestation of the holographic principle [173,

169, 27]—gauge invariant degrees of freedom in a quantum theory of gravity live in some "holographic space."

- Despite the strong reference frame dependence of the location of the microscopic degrees of freedom, the answers to any physical question are consistent with each other when asked in different reference frames. In particular, when we change the reference frame, the distribution of the microscopic degrees of freedom (as well as some of the semiclassical degrees of freedom) is rearranged such that this consistency is maintained.

These items are basic features of general coordinate transformations at the level of full quantum gravity, beyond the approximation of semiclassical theory in curved spacetime. In particular, they provide important clues about how complementarity as envisioned in Refs. [133, 140] may be realized at the microscopic level.

2.5 Summary—A Grand Picture

The relation between the quantum mechanical view of the world and the spacetime picture of general relativity has never been clear. The issue becomes particularly prominent in a system with a black hole. Quantum mechanics suggests that the black hole formation and evaporation processes are unitary—a black hole appears simply as an intermediate (gigantic) resonance between the initial collapsing matter and final Hawking radiation states. On the other hand, general relativity suggests that a classical observer falling into a large black hole does not feel anything special at the horizon. These two, seemingly unrelated, assertions are surprisingly hard to reconcile. With naive applications of standard quantum field theory on curved spacetime, one is led to the conclusion that unitarity of quantum mechanics is violated [85] or that an infalling observer finds something dramatic (firewalls) at the location of the horizon [10, 9, 126, 38, 127].

In this paper, we have argued that the resolution to this puzzle lies in how a semiclassical description of the system—quantum theory of matter and radiation on a fixed spacetime background—arises from the microscopic theory of quantum gravity. While a semiclassical description employs an *exact* spacetime background, the quantum uncertainty principle implies that there is no such thing—there is an intrinsic uncertainty for background spacetime for any finite energy and momentum. This implies, in particular, that at the microscopic level there are many different ways to arrive at the same background for the semiclassical theory, within the precision allowed by quantum mechanics. This is the origin of the Bekenstein-Hawking (and related, e.g. Gibbons-Hawking [70]) entropy. The semiclassical picture is obtained after coarse-graining these degrees of freedom representing the microscopic structure of spacetime, which we called the vacuum degrees of freedom. More specifically, any result in semiclassical theory is a statement about the maximally mixed ensemble of microscopic quantum states consistent with the specified background within the required uncertainty [141].

This picture elucidates why the purely semiclassical calculation of Ref. [85] finds a violation of unitarity. At the microscopic level, formation and evaporation of a black hole are processes in which information in the initial collapsing matter is converted into that in the vacuum degrees of freedom, which is later transferred back to semiclassical degrees of freedom, i.e. Hawking radiation. Since semiclassical theory is incapable of describing microscopic details of the vacuum degrees of freedom (because it describes them as already coarse-grained, Bekenstein-Hawking entropy), the *description* of the black hole formation and evaporation processes in semiclassical theory violates unitarity at all stages throughout these processes. This, of course, does not mean that the processes are non-unitary at the fundamental level.

In order to address the unitary evolution and explore its relation with the existence or non-existence of the interior spacetime, we therefore need to discuss the properties of the vacuum degrees of freedom. While the theory governing the detailed microscopic dynamics of these degrees of freedom is not yet fully known, we may include them in our description in the form of a new index—vacuum index—carried by the microscopic quantum states (which we denoted by k and κ) in addition to the indices representing excitations in semiclassical theory and of the stretched horizon. We have argued that these degrees of freedom show peculiar features, which play key roles in addressing the paradoxes discussed in Refs. [10, 9, 126]:

Extreme relativeness:

In a fixed reference frame, vacuum degrees of freedom may be viewed as distributed (nonlocally) over space. The spacetime distribution of these degrees of freedom, however, changes if we adopt a different reference frame—they are not anchored to spacetime, and rather live in some "holographic space." This dependence on the reference frame occurs in a way that the answers to any physical question are consistent with each other when asked in different reference frames. Together with the reference frame dependence of (some of the) semiclassical degrees of freedom, discussed in the earlier literature [170, 166], this comprises basic features of how general coordinate transformations work in the full theory of quantum gravity.

Spacetime-matter duality:

The vacuum degrees of freedom exhibit dual properties of spacetime and matter (even in a description in a single reference frame): while these degrees of freedom are interpreted as representing the way the semiclassical spacetime is realized at the microscopic level, their interactions with semiclassical degrees of freedom make them look like thermal radiation. (At a technical level, the Hilbert space labeled by the vacuum index and that by semiclassical excitations do not factor.) In a sense, these degrees of freedom are neither spacetime nor matter/radiation, as can be seen from the fact that their spacetime distribution changes as we change the reference frame, and that their detailed dynamics cannot be treated in semiclassical theory (as was done in Refs. [10,

9, 126]). This situation reminds us of wave-particle duality, which played an important role in early days in the development of quantum mechanics—a quantum object exhibited dual properties of waves and particles, while the "true" (quantum) description did not fundamentally rely on either of these classical concepts.

These features make the existence of the black hole interior consistent with unitary evolution, in the sense of complementarity [170] as envisioned in Refs. [133, 140]. In particular, a large nearly flat spacetime region near the Schwarzschild horizon becomes manifest in a reference frame whose origin follows a free-fall trajectory starting from rest from a point sufficiently far from the black hole.

It is often assumed that two systems related by the equivalence principle, e.g. a static detector held near the Schwarzschild horizon and an accelerating detector in Minkowski space, must reveal similar physics. This is, however, not true. Since the equivalence principle can make a statement only about a point at a given moment in a given reference frame, while a system in quantum mechanics is specified by a state which generally encodes global information on the equal-time hypersurface, there is no reason that physics of the two systems must be similar beyond a point in space. In particular, a detector reacts very differently to blueshifted Hawking radiation and Unruh radiation in Minkowski space at the microscopic level—it extracts microscopic information about spacetime in the former case, while it does not in the latter.

While our study has focused on a system with a black hole, we do not see any reason why the basic picture we arrived at does not apply to more general cases. We find it enlightening that our results indicate specific properties for the microscopic degrees of freedom that play a crucial role in the emergence of spacetime at the fundamental level. Unraveling the detailed dynamics of these degrees of freedom would be a major step toward obtaining a complete theory of quantum gravity. As a first step, it seems interesting to study implications of our picture for the case that spacetime approaches anti-de Sitter space in the asymptotic region, in which we seem to know a little more [122]. It would also be interesting to explore implications of our picture for cosmology, e.g. along the lines of Refs. [133, 132, 134].

Chapter 3

The Black Hole Interior in Quantum Gravity

3.1 Introduction

Despite much effort, the relation between quantum mechanics and the spacetime picture of general relativity has never been clear. The issue becomes particularly prominent in black hole physics [154]. Quantum mechanics suggests that the black hole formation and evaporation processes are unitary—a black hole simply appears as an intermediate resonance between the initial collapsing matter and final Hawking radiation states [174]. Meanwhile, general relativity suggests that an observer falling into a large black hole does not feel anything special at the horizon. These two assertions are surprisingly hard to reconcile. With naive applications of quantum field theory on curved spacetime, one is led to the conclusion that unitarity of quantum mechanics is violated [85] or an infalling observer finds something dramatic (a firewall) at the horizon [10, 9, 126, 38].

In this letter, we argue that the resolution to this puzzle lies in how a semiclassical description of the system arises from the microscopic theory of quantum gravity. While a semiclassical description employs an *exact* spacetime background, the quantum uncertainty principle implies that there is no such thing—there is an intrinsic uncertainty for background spacetime for any finite energy and momentum. This implies that at the microscopic level there are many different ways to arrive at the same background for the semiclassical theory, within the precision allowed by quantum mechanics. This is the origin of the Bekenstein-Hawking entropy [22, 86]. The semiclassical picture is obtained after coarse-graining these degrees of freedom, which we call *vacuum degrees of freedom* [141].

We argue that much of the puzzle regarding unitary evolution and the interior spacetime of a black hole arises from peculiar features the vacuum degrees of freedom exhibit when viewed from the semiclassical standpoint. In particular, they show properties which we call *extreme relativeness* and *spacetime-matter duality*. The first refers to the fact that the spacetime distribution of these degrees of freedom changes when we adopt a different

"reference frame." This change occurs in a way that the answers to any physical question are consistent with each other when asked in different reference frames. Together with the reference frame dependence of the semiclassical degrees of freedom discussed earlier [170, 166], this comprises basic features of how general coordinate transformations work in the full theory of quantum gravity.

The second property is related to the following fact: while the vacuum degrees of freedom are interpreted as how the semiclassical spacetime is realized at the microscopic level, their interactions with semiclassical degrees of freedom make them look like thermal radiation. In fact, these degrees of freedom are neither spacetime nor matter/radiation, as indicated by the fact that their spacetime distribution is frame dependent, and that their detailed dynamics cannot be treated in semiclassical theory. This situation reminds us of wave-particle duality—a quantum object exhibits dual properties of waves and particles while the "true" (quantum) description does not fundamentally rely on either of these classical concepts.

The two properties described above allow us to avoid the arguments in Refs. [10, 9, 126] and make the existence of the black hole interior consistent with unitary evolution, in the sense of complementarity [170] as envisioned in Refs. [132, 140]. A notion of geometry carrying information has also been considered recently in Ref. [152] in a different model of black hole evolution; see also Ref. [155] for early discussions. In our picture, we assume that a black hole evaporates through Hawking radiation [86]; for an alternative view, see Ref. [79].

In the rest of the letter, we present our picture using the example of a Schwarzschild black hole formed by collapsing matter in 4-dimensional spacetime. More detailed descriptions are given in the accompanying paper [137].

3.2 Distant Description

Consider a quantum state representing a black hole of mass M located at some place at rest, as described in a distant reference frame. (We adopt the Schrödinger picture throughout.) Because of the uncertainty principle, such a state must involve a superposition of energy and momentum eigenstates. In particular, since a black hole of mass M will evolve after Schwarzschild time $\Delta t \approx O(M l_{\rm P}^2)$ into a state representing a Hawking quantum and a smaller mass black hole, the state must involve a superposition with

$$\Delta E \approx \frac{1}{\Delta t} \approx O\left(\frac{1}{M l_{\rm P}^2}\right), \tag{3.1}$$

where E is defined in the asymptotic region, and $l_{\rm P}$ the Planck length. Requiring that the position uncertainty is comparable to the quantum stretching of the horizon $\Delta r \approx O(1/M)$, where r is the Schwarzschild radial coordinate, the momentum spread is $\Delta p \approx O(1/M l_{\rm P}^2)$. This gives an uncertainty of the kinetic energy much smaller than ΔE, so the spread of the energy comes mostly from a superposition of different rest masses: $\Delta E \approx \Delta M$.

How many different independent ways are there to superpose the energy eigenstates to arrive at the same black hole geometry within this precision? We assume that the Bekenstein-Hawking entropy, $\mathcal{A}/4l_{\rm P}^2$, gives the logarithm of this number (at the leading order in $l_{\rm P}^2/\mathcal{A}$), where $\mathcal{A} = 16\pi M^2 l_{\rm P}^4$ is the area of the horizon. The nonzero Bekenstein-Hawking entropy implies that there are exponentially many independent black hole *vacuum* states in a small energy interval of Eq. (3.1):

$$S_0 = \frac{\mathcal{A}}{4l_{\rm P}^2} + O\left(\frac{\mathcal{A}^q}{l_{\rm P}^{2q}};\, q < 1\right), \tag{3.2}$$

i.e. the states that do not have a field/string theoretic excitation on the semiclassical black hole background and in which the stretched horizon, located at $r = 2Ml_{\rm P}^2 + O(1/M) \equiv r_{\rm s}$, is not excited.

Labeling these exponentially many states by k, which we call the *vacuum index*, basis states for the general microstates of a black hole of mass M (within the uncertainty ΔM) can be given by

$$|\Psi_{\bar{a}\, a\, a_{\rm far};k}(M)\rangle \approx |\psi_{\bar{a}a;k}(M)\rangle |\phi_{a_{\rm far}}(M)\rangle. \tag{3.3}$$

Here, $\bar{a}$, a, and $a_{\rm far}$ label the excitations of the stretched horizon, in the zone (i.e. the region within the gravitational potential barrier defined, e.g., as $r \leq R_{\rm Z} \equiv 3Ml_{\rm P}^2$), and outside the zone ($r > R_{\rm Z}$), respectively, and $|\psi_{\bar{a}a;k}(M)\rangle$ and $|\phi_{a_{\rm far}}(M)\rangle$ are black hole and exterior states. (Here, we have used the fact that k can be regarded as being mostly in $r \leq R_{\rm Z}$; see later.) As we have argued, the index k runs over $1, \cdots, e^{S_0}$ for the vacuum states $\bar{a} = a = a_{\rm far} = 0$. In general, the range for k depends on $\bar{a}$ and a, but its dependence is higher order in $l_{\rm P}^2/\mathcal{A}$ so we mostly ignore it. This small dependence, however, becomes relevant when we discuss negative energy excitations associated with Hawking emission.

Excitations here are defined as fluctuations with respect to a fixed background, so their energies as well as entropies can be either positive or negative, although their signs must be the same. As discussed in Refs. [173, 142], the contribution of the excitations to the total entropy is subdominant in $l_{\rm P}^2/\mathcal{A}$. The total entropy in the near black hole region, $r \leq R_{\rm Z}$, is thus given by $S = \mathcal{A}/4l_{\rm P}^2$ at the leading order.

The fact that all the independent microstates with different k lead to the same geometry suggests that the semiclassical picture is obtained after coarse-graining the degrees of freedom represented by this index, the vacuum degrees of freedom [141]. According to this picture, the black hole vacuum state in the semiclassical description is given by the density matrix

$$\rho_0(M) = \frac{1}{e^{S_0}} \sum_{k=1}^{e^{S_0}} |\Psi_{\bar{a}=a=a_{\rm far}=0;k}(M)\rangle\langle\Psi_{\bar{a}=a=a_{\rm far}=0;k}(M)|. \tag{3.4}$$

To obtain the response of this state to the operators in the semiclassical theory, we may trace out the subsystem on which they do not act. Denoting this subsystem by $\bar{C}$, the relevant reduced density matrix is

$$\tilde{\rho}_0(M) = {\rm Tr}_{\bar{C}}\, \rho_0(M). \tag{3.5}$$

Consistently with our identification of the origin of the Bekenstein-Hawking entropy, we assume that this represents the thermal density matrix

$$\tilde{\rho}_0(M) \approx \frac{e^{-\beta H_{\rm sc}(M)}}{{\rm Tr}\, e^{-\beta H_{\rm sc}(M)}}; \quad \beta = \begin{cases} \frac{1}{T_{\rm H}} & \text{for } r \le R_{\rm Z}, \\ +\infty & \text{for } r > R_{\rm Z}, \end{cases} \tag{3.6}$$

where $T_{\rm H} = 1/8\pi M l_{\rm P}^2$, and $H_{\rm sc}(M)$ is the Hamiltonian of the semiclassical theory.

In standard semiclassical field theory, the density matrix of Eq. (3.6) is obtained as a reduced density matrix by tracing out the region within the horizon in the *unique* global black hole vacuum state. Our view is that this density matrix is obtained from a mixed state of exponentially many pure states, arising from the coarse-graining in Eq. (3.4). We stress that the information in the vacuum index k is invisible in the semiclassical theory as it is already coarse-grained *to obtain* the theory; in particular, the dynamics of the vacuum degrees of freedom cannot be described in terms of $H_{\rm sc}(M)$.

The expression in Eq. (3.6) suggests that the spatial distribution of the information about k follows the thermal entropy calculated using the local temperature:

$$T(r) \simeq \begin{cases} \frac{T_{\rm H}}{\sqrt{1-\frac{2Ml_{\rm P}^2}{r}}} & \text{for } r \le R_{\rm Z}, \\ 0 & \text{for } r > R_{\rm Z}. \end{cases} \tag{3.7}$$

In particular, the region around the edge of the zone, $r \le R_{\rm Z}$ and $r - 2Ml_{\rm P}^2 \not\ll Ml_{\rm P}^2$, contains $O(1)$ bits of information about k.

Semiclassical operators in the zone act nontrivially on both a and k indices; otherwise the maximal mixture in Eq. (3.4) is not compatible with the thermality in Eq. (3.6). Since the thermal nature is prominent only for modes whose energies measured in the asymptotic region are

$$\omega \lesssim T_{\rm H}, \tag{3.8}$$

this feature is significant only for such infrared modes. For operators with Eq. (3.8), their actions on microstates can be complicated, although they act on the coarse-grained vacuum state of Eq. (3.4) as if it is the thermal state in Eq. (3.6).

There is a simple physical picture behind this phenomenon of "non-decoupling" of the a and k indices for the infrared modes. As viewed from a distance, these modes are "too soft" to be resolved clearly above the background. Since the derivation of the semiclassical theory involves coarse-graining over microstates in which the energy stored in the region $r \lesssim R_{\rm Z}$ has spreads of order $\Delta E \approx 1/Ml_{\rm P}^2$, infrared modes with $\omega \lesssim T_{\rm H} \approx O(1/Ml_{\rm P}^2)$ are not necessarily distinguished from "spacetime fluctuations" of order ΔE.

The structure described above leads to the following picture for black hole evaporation.[1] Suppose a black hole of mass M is in microstate k:

$$|\Psi_k(M)\rangle = |\psi_k(M)\rangle|\phi_I\rangle, \tag{3.9}$$

[1] We focus on a single Hawking emission and ignore excitations beyond those directly associated with the emission. For a more complete discussion, see Ref. [137].

where $|\psi_k(M)\rangle$ is the black hole state, with suppressed excitation indices, and $|\phi_I\rangle$ the exterior state. After a timescale of $t \approx O(Ml_{\rm P}^2)$, this state evolves due to Hawking emission as

$$|\psi_k(M)\rangle|\phi_I\rangle \to \sum_{i,a,k'} c^k_{iak'}|\psi_{a;k'}(M)\rangle|\phi_{I+i}\rangle, \tag{3.10}$$

where $|\phi_{I+i}\rangle$ is the state in which newly emitted Hawking quanta, labeled by i and having energy E_i, are added to the appropriately time evolved $|\phi_I\rangle$. The index a represents the fact that the black hole state has negative energy excitations of energy $-E_a$ around the edge of the zone, created in connection with the Hawking emission; the coefficients $c^k_{iak'}$ are nonzero only if $E_i \approx E_a$ (within the uncertainty). The negative energy excitations then propagate inward, and after a time of order $Ml_{\rm P}^2 \ln(Ml_{\rm P})$ collide with the stretched horizon, making the black hole states relax as

$$|\psi_{a;k'}(M)\rangle \to \sum_{k_a} d^{ak'}_{k_a}|\psi_{k_a}(M-E_a)\rangle. \tag{3.11}$$

The combination of Eqs. (3.10, 3.11) yields

$$|\psi_k(M)\rangle|\phi_I\rangle \to \sum_{i,k_i} \alpha^k_{ik_i}|\psi_{k_i}(M-E_i)\rangle|\phi_{I+i}\rangle, \tag{3.12}$$

where $\alpha^k_{ik_i} = \sum_{a,k'} c^k_{iak'} d^{ak'}_{k_i}$, and we have used $E_i = E_a$. This expression shows that information in the black hole can be transferred to the radiation state i.

It is important that the negative energy excitations in Eq. (3.10) come with *negative entropies*, so that each of the processes in Eqs. (3.10, 3.11) is separately unitary. Specifically, as k and i run over all the possible values with a being fixed, the index k' runs only over $1, \cdots, e^{S_0(M-E_a)}$, the dimension of the space spanned by k_a. This is an example of the non-factorizable nature of the k and a indices discussed after Eq. (3.3). This structure avoids the firewall argument in Ref. [9]—unlike what is imagined there, the physical Hilbert space is smaller than the naive Fock space built on each k.

From the semiclassical standpoint, the emission of Eq. (3.10) is viewed as occurring locally around the edge of the zone, which is possible because the information about the black hole microstate extends into the whole zone region. In this region, information stored in the vacuum state, k, is transferred into that in modes $a_{\rm far} \neq 0$, which have clear identities over the background spacetime. Due to energy conservation, this process is accompanied by the creation of ingoing negative energy excitations, which are *not* entangled with the emitted Hawking quanta.

The discussion here indicates that the purifiers of the emitted Hawking quanta are microstates which semiclassical theory describes as a vacuum. Unlike what was considered in Ref. [10], Hawking quanta are not modes associated solely with one of the Rindler wedges in the near horizon approximation (b modes in the notation of Ref. [10]) nor outgoing Minkowski modes (a modes), which would appear to have high energies for infalling observers. This allows for avoiding the entropy [10] and typicality [126] arguments for firewalls. Note that

physics described here need not introduce nonlocality in low energy field theory; it can still respect causality in $r > r_{\rm s}$.

We emphasize that the vacuum degrees of freedom play *dual* roles. While they represent how the semiclassical spacetime is composed at the microscopic level, they also appear as thermal radiation when probed in the semiclassical theory. In fact, these degrees of freedom are neither spacetime nor matter/radiation. In particular, their detailed dynamics cannot be treated in semiclassical theory.

The above understanding of Hawking emission clarifies why the semiclassical calculation of Ref. [85] finds an apparent violation of unitarity. At the microscopic level, formation and evaporation of a black hole involve the vacuum degrees of freedom. Since semiclassical theory is incapable of describing their microscopic dynamics, the description of black hole evolution in semiclassical theory is necessarily non-unitary.

A similar analysis can also be performed for black hole mining [180, 39]. See Ref. [137] for details.

3.3 Infalling Description

Suppose we drop an object into a black hole. In a distant reference frame, the semiclassical description of the object (in terms of a and $a_{\rm far}$) is applicable only until it hits the stretched horizon, after which it is represented as excitations of the stretched horizon (in terms of $\bar{a}$). The information about the fallen object will then stay there, at least, for the scrambling time of order $Ml_{\rm P}^2 \ln(Ml_{\rm P})$ [88] before being transferred to k. On the other hand, the equivalence principle says that the falling object does not feel anything special when it crosses the horizon. How can these two pictures be consistent?

The idea of complementarity is that the infalling object is still described using low energy language after it crosses the Schwarzschild horizon by making an appropriate reference frame change. Here we consider a class of reference frames which reveal the spacetime structure near the Schwarzschild horizon in the clearest form. We call them *infalling reference frames*.

Let the spatial origin p_0 of a reference frame follow a timelike geodesic released from rest at $r = r_0$, with $r_0 - 2Ml_{\rm P}^2 \gtrsim Ml_{\rm P}^2$. According to complementarity, the system described in this reference frame does not have a (hot) stretched horizon at the location of the Schwarzschild horizon when p_0 crosses it; the region around p_0 appears approximately flat up to small curvature effects.

In this description, a "horizon" signaling the breakdown of the semiclassical description is expected to appear in the past-directed and inward directions from p_0. In analogy with the case of a distant frame description, we denote basis states for the general microstates as

$$|\Psi_{\bar{\alpha}\,\alpha\,\alpha_{\rm far};\kappa}(M)\rangle, \tag{3.13}$$

where $\bar{\alpha}$ labels the excitations of the "horizon," and α, and $\alpha_{\rm far}$ the semiclassical excitations near and far from the black hole, respectively; κ is the vacuum index.

The complementarity transformation provides a map between the states in Eq. (3.3) and those in Eq. (3.13). While the general form of this transformation can be complicated, we may consider, based on the analysis of an infalling object, that a portion of the α index representing interior excitations is transformed into the $\bar{a}$ index (and vice versa). Note that the amount of information needed to reconstruct the interior (in the semiclassical sense) is much smaller than the Bekenstein-Hawking entropy—the logarithm of the dimension of the relevant Hilbert space is of order $(\mathcal{A}/l_{\rm P}^2)^q$ with $q < 1$.

Where are the κ degrees of freedom located? We expect that most are in the region close to the "horizon"; in particular, the number of κ degrees of freedom within a distance sufficiently smaller than $Ml_{\rm P}^2$ from p_0 is of $O(1)$, since the time and length scales characterizing local deviations from Minkowski space are of order $Ml_{\rm P}^2$ there. This invites a question: how can this picture be consistent with that in the distant reference frame, which has a very different spacetime distribution of the vacuum degrees of freedom?

To see a nontrivial consistency between the two pictures, consider detectors hovering at a constant r with $r - 2Ml_{\rm P}^2 \ll Ml_{\rm P}^2$. In a distant description, the spatial density of the microscopic information in k is large there, so that these detectors can be used for black hole mining. The rate of extracting information, however, is still of order one qubit per Schwarzschild time $t \approx O(Ml_{\rm P}^2)$ *per channel* [39]—the acceleration of information extraction occurs not because of a higher rate in each channel but because of an increased number of available channels. This implies that each single detector, which we define to act on a single channel, "clicks" once per $t \approx O(Ml_{\rm P}^2)$.

In an infalling reference frame, the density of the microscopic information in κ is small at the detector location, at least when p_0 passes nearby. The rate of extracting information thus cannot be much faster than $1/Ml_{\rm P}^2$ around p_0, reflecting the fact that the spacetime appears approximately flat there. This, however, is still consistent with the distant description. By adopting the near-horizon Rindler approximation, one can show that when viewed from the infalling reference frame, the detector clicks only once in each time/space interval of

$$\Delta T \approx \Delta Z \approx O(Ml_{\rm P}^2), \tag{3.14}$$

around p_0 [137]. This is what we expect from the equivalence principle: the spacetime appears flat up to curvature effects with lengthscale $Ml_{\rm P}^2$. While the detector clicks of order $\ln(Ml_{\rm P})$ times within the causal patch of the infalling frame, these clicks occur at distances of order $Ml_{\rm P}^2$ away from p_0, where we expect a higher density of κ degrees of freedom.

The two descriptions are thus consistent. It is striking that the microscopic information about a black hole exhibits this level of reference frame dependence, a phenomenon we refer to as *extreme relativeness*.

3.4 Other Reference Frames

We now discuss a reference frame whose origin follows a timelike geodesic released from rest at $r = r_0$, where r_0 is *close to* the Schwarzschild horizon, $r_0 - 2Ml_{\rm P}^2 \ll Ml_{\rm P}^2$. In the case of

$r_0 - 2Ml_{\rm P}^2 \gtrsim Ml_{\rm P}^2$, we found that the detector-click time/length scales are given by Eq. (3.14), despite the fact that the detector clicks at a much higher rate in its own frame. Technically, this was due to a huge relative boost between p_0 and the detector when they approach. Here, however, the relevant boost is not as large, and the detector-click time/length scales appear as

$$\Delta T \approx \Delta Z \ll Ml_{\rm P}^2. \tag{3.15}$$

Since each detector click extracts an $O(1)$ amount of information from spacetime, which we expect not to occur in Minkowski space, this implies that the spacetime as viewed from this reference frame is not approximately Minkowski over the lengthscale $Ml_{\rm P}^2$ when p_0 crosses the Schwarzschild horizon. We interpret this to mean that in this reference frame, the "horizon" is at a distance of order ΔZ away from p_0, so that detector clicks occur near or "on" this surface. Since we expect that the microscopic information is located near and on the "horizon," there is no inconsistency for the clicks to extract information from the black hole.

One might worry that in this reference frame, spacetime near the Schwarzschild horizon does not appear large, $\approx O(Ml_{\rm P}^2)$, nearly flat space. However, the *existence* of an infalling reference frame discussed before ensures that an infalling physical observer sees a large black hole interior. The analysis here simply says that the spacetime around the Schwarzschild horizon is not always *described* as a large nearly flat region, even in reference frames falling freely into the black hole.

We finally discuss (non-)relations of black hole mining and the Unruh effect [179] in Minkowski space. It is often thought that these two reveal the same physics, which would mean the existence of a "horizon" in an *inertial* frame description of Minkowski space. This is, however, not true. Since the equivalence principle can make a statement only about a point at a given moment in a given reference frame, while a system in quantum mechanics is specified by a state which encodes global information on the equal-time hypersurface, there is no reason that physics of the two systems must be similar beyond a point in space. In particular, the inertial frame description of Minkowski space does not have a "horizon," so a detector reacts very differently to blueshifted Hawking radiation and Unruh radiation in Minkowski space—it extracts microscopic information about spacetime in the former case, while it does not in the latter. The relation between quantum mechanics and the equivalence principle seems subtle, but they are consistent.

Chapter 4

Axion Isocurvature and Magnetic Monopoles

4.1 Introduction

Cosmic inflation not only provides a framework to address many puzzles of early universe cosmology [75, 115, 6] but also incorporates a mechanism that seeds the formation of the structure in the universe [87, 164, 77]. An exciting aspect of the inflationary mechanism is that it also sources gravitational waves. If inflation occurs at a sufficiently high scale ($\sim 10^{15}$–10^{16} GeV), the amplitude of these gravitational waves is large enough to leave a measurable imprint on the polarization of the cosmic microwave background (CMB) [192, 105]. A number of CMB polarization experiments are presently searching for this signal [1, 112]. A positive signal in such an experiment would have interesting implications for particle physics, especially for ultra-light bosonic fields. Bosonic fields with masses lighter than the inflationary Hubble scale are efficiently produced by inflation and can cause isocurvature perturbations in the CMB [118, 119, 160, 178]. High scale inflation thus leads to interesting constraints on ultra-light bosons, including the QCD axion provided the axion decay constant f_a is greater than the inflationary scale.

It is widely regarded [67, 23, 96, 121] that a discovery of inflationary gravitational waves would rule out the QCD axion with a decay constant $f_a \gtrsim 10^{16}$ GeV, a range that is favored by several theoretical considerations [171, 11]. Experiments have also been proposed recently to search for the QCD axion in this parameter range [72, 73, 40], and it is of great interest to delineate the viable parameter space accessible to these efforts. For example, this bound disappears if the QCD axion acquires a large mass during inflation, damping the production of isocurvature modes. At the end of inflation, however, this mass has to nearly vanish for the QCD axion to solve the strong CP problem. While models achieving this do exist (see [54, 175] for example), they face the difficulty that the mechanism responsible for generating a large axion mass during inflation has to violate the Peccei-Quinn symmetry while ensuring that this violation remains sufficiently sequestered from the axion after inflation. This task

is made even more difficult by the fact that these dynamics must couple to the inflaton. Other proposals to alleviate the tension between high scale inflation and the QCD axion include a dynamically changing Peccei-Quinn breaking scale [108]. While reasonable, such models sacrifice some of the theoretical arguments underlying high f_a axions. There are also attempts that involve transfer of the axion isocurvature from one species to another [111], but these typically deplete the dark matter abundance of the axion, eliminating one of the promising ways to search for them. It might also be possible to relax these constraints by dumping entropy into the universe around the QCD phase transition [109], but these channels are constrained [67, 23, 96, 121].

In this paper, we investigate an alternative possibility: what if the QCD axion acquires a large mass *after* inflation, which subsequently disappears *before* the QCD phase transition? If this mass is larger than the Hubble scale during a large interval, somewhere between the reheating and QCD scales, then the axion field oscillates earlier and the fluctuations in the field will be damped, relaxing into the minimum of the potential generating this large mass. When this mass (and potential) subsequently disappears, the average axion field takes a value corresponding to this minimum. Since this minimum is in general displaced from the QCD minimum, the misalignment between these two points regenerate a cosmic abundance of the QCD axion when the axion reacquires a mass during the QCD phase transition, enabling it to be dark matter. The isocurvature perturbations, however, will be small since the initial evolution of the field causes the perturbations to coalesce around the initial minimum, while the subsequent dark matter abundance is generated by the homogeneous misalignment between the QCD minimum and the initial minimum.

How can we give such a large initial mass that then disappears almost completely? We accomplish this by coupling the QCD axion to a new $U(1)'$ gauge group. If the reheating of the universe produces magnetic monopoles under this $U(1)'$, the monopole density generates a mass for the axion [66]. This is because topological terms like $F\tilde{F}$ become physical in the presence of magnetic monopoles due to the Witten effect [190]. Specifically, it gives a free energy density that depends on a background axion field value, thus creating an effective mass for the axion. This mass is sufficient to damp isocurvature perturbations in the axion field. After the perturbations have been damped, the monopole density can be efficiently eliminated by breaking the $U(1)'$ symmetry, resulting in confinement and subsequent annihilation of the monopoles. The monopole density forces the axion field to relax into θ', a point on the potential chosen by CP phases in the $U(1)'$ sector. Since this phase need not be aligned with the QCD minimum at $\theta_{\rm QCD}$, the axion generally acquires a homogeneous cosmic abundance during the QCD phase transition, with suppressed isocurvature perturbations. For large $f_a \gg 10^{12}$ GeV, this misalignment needs to be small, $|\theta' - \theta_{\rm QCD}| \ll 1$, but this can be environmentally selected [120, 177]. We show that there is sufficient time for the damping of axion isocurvature fluctuations so that axion dark matter with a unification scale f_a is consistent with high scale inflation giving an observable size of the gravitational wave polarization signal.

The organization of this paper is as follows. In Section 4.2, we review the required amount of damping of axion isocurvature fluctuations consistent with current observations.

In Section 4.3 we introduce our basic mechanism, and in Section 4.4 we present a minimal model realizing it. We show that the model can consistently accommodate unification scale axion dark matter with high scale (unification scale) inflation. In Section 4.5, we discuss monopole annihilations due to $U(1)'$ breaking in detail, showing that they efficiently eliminate monopoles. In Section 4.6, we discuss extensions/modifications of the minimal model in which the issue of radiative stability of the $U(1)'$ sector existing in the minimal model does not arise. We conclude in Section 4.7.

4.2 Required Damping of Isocurvature Perturbations

Inflation generally induces quantum fluctuations of order $H_{\rm inf}/2\pi$ for any massless field, where $H_{\rm inf}$ is the Hubble parameter during inflation. This implies that if $U(1)_{\rm PQ}$ is broken before or during inflation, then the angle $\theta = a/f_a$ of the axion field a has fluctuations

$$\delta\theta(T_{\rm R}) \approx \frac{H_{\rm inf}}{2\pi f_a}, \tag{4.1}$$

at temperature $T_{\rm R}$, when the radiation dominated era starts due to reheating.[1] Since the axion potential is flat during inflation, these fluctuations are converted to isocurvature density perturbations upon the generation of the axion mass.

There is a tight constraint on the amount of allowed isocurvature perturbations from the Planck data [2, 3], which can be written as (see, e.g., [43])

$$\frac{\Omega_a}{\Omega_{\rm DM}} \frac{\delta\theta(T_{\rm QCD})}{\theta_{\rm mis}} \lesssim 4.8 \times 10^{-6}, \tag{4.2}$$

where $\theta_{\rm mis}$ is the average axion misalignment angle, while $\delta\theta(T_{\rm QCD})$ is the angle fluctuation of the axion field at temperature $T_{\rm QCD} \sim 1$ GeV. Here, Ω_a and $\Omega_{\rm DM} \simeq 0.24$ represent the axion and total dark matter abundances, respectively, and we assume $\theta_{\rm mis} > \delta\theta(T)$ throughout.[2] Using the expression for the axion relic density

$$\frac{\Omega_a}{\Omega_{DM}} \approx 1.0 \times 10^5\, \theta_{\rm mis}^2 \left(\frac{f_a}{10^{16}\ {\rm GeV}}\right)^{1.19}, \tag{4.3}$$

(which requires $\theta_{\rm mis} \lesssim 0.003$ for $f_a \simeq 10^{16}$ GeV, possibly realized through environmental selection effects [120, 177]), we may rewrite Eq. (4.2) as

$$\delta\theta(T_{\rm QCD}) \lesssim 1.5 \times 10^{-8} \sqrt{\frac{\Omega_{\rm DM}}{\Omega_a}} \left(\frac{10^{16}\ {\rm GeV}}{f_a}\right)^{0.6}. \tag{4.4}$$

[1] In this paper we adopt the instant reheating approximation for simplicity, so that the universe is radiation dominated right after inflation. An extension of our analysis to more general cases (including a matter dominated era before reheating) is straightforward.

[2] This condition requires $H_{\rm inf} \lesssim 2 \times 10^{14}\ {\rm GeV} \sqrt{\Omega_a/\Omega_{\rm DM}} (f_a/10^{16}\ {\rm GeV})^{0.4}$; for comparison, see Eq. (4.5) and an estimate below it for unification scale inflation.

Assuming the standard cosmological history after inflation, $\delta\theta(T_{\rm QCD}) \approx \delta\theta(T_{\rm R})$, so that we find

$$H_{\rm inf} \lesssim 9.4 \times 10^8 \text{ GeV} \sqrt{\frac{\Omega_{\rm DM}}{\Omega_a}} \left(\frac{f_a}{10^{16} \text{ GeV}} \right)^{0.4}. \tag{4.5}$$

This severely constrains inflationary models in the presence of a unification scale axion [67, 23, 96, 121]. In particular, unification scale axion dark matter—$\Omega_a = \Omega_{\rm DM}$ and $f_a \sim 10^{16}$ GeV—is inconsistent with unification scale inflation—$E_{\rm inf} \equiv V_{\rm inf}^{1/4} \sim 10^{16}$ GeV, which leads to $H_{\rm inf} = E_{\rm inf}^2/\sqrt{3}\bar{M}_{\rm Pl} \sim 10^{13}$ GeV, where $\bar{M}_{\rm Pl} \simeq 2.4 \times 10^{18}$ GeV is the reduced Planck scale.

Below, we discuss a scenario in which axion isocurvature fluctuations are damped due to dynamics after inflation. Defining the (inverse) damping factor Δ by

$$\Delta = \frac{\delta\theta(T_{\rm QCD})}{\delta\theta(T_{\rm R})}, \tag{4.6}$$

Eq. (4.4) yields

$$\Delta \lesssim 1 \times 10^{-4} \sqrt{\frac{\Omega_{\rm DM}}{\Omega_a}} \left(\frac{f_a}{10^{16} \text{ GeV}} \right)^{0.4} \left(\frac{10^{13} \text{ GeV}}{H_{\rm inf}} \right). \tag{4.7}$$

Here, we have normalized f_a and $H_{\rm inf}$ by the values corresponding to unification scale axion and inflation, respectively. This gives the required amount of damping.

4.3 Basic Mechanism

Our basic idea of suppressing axion isocurvature fluctuations is that the axion mass obtains extra contributions beyond that from QCD in the early universe so that it is larger than the Hubble parameter in some period. In this period, axion isocurvature perturbations are reduced because of the damped oscillations of the axion field, giving $\Delta < 1$.

We do this by introducing a coupling of the axion to a hidden $U(1)'$ gauge group

$$\mathcal{L} \sim \frac{1}{f_a} a F'^{\mu\nu} \tilde{F}'_{\mu\nu}. \tag{4.8}$$

We assume that at some temperature T_M after inflation ($T_M \lesssim T_{\rm R} \approx E_{\rm inf}$), monopoles of $U(1)'$ are created. This can happen, for example, if a hidden sector $SU(2)'$ gauge group is broken to $U(1)'$ at that scale.[3] In the presence of magnetic monopoles, the coupling in Eq. (4.8) induces an effective mass for the axion [66]:

$$m_a^2(T) = \gamma \frac{n_M(T)}{f_a}, \tag{4.9}$$

[3] If the creation of monopoles is associated with $G \to G' \times U(1)'$ symmetry breaking in the hidden sector, where G and G' are non-Abelian gauge groups, then we would need to have two axion fields in the ultraviolet so that the QCD axion remains after G' gives a large mass to one linear combination of the two axion fields.

where γ is determined by the structure of the $U(1)'$ sector, such as the gauge coupling and matter content. (γ may in general depend on temperature, although it is not the case in the explicit model considered below.) $n_M(T)$ is the number density of the monopoles; assuming the abundance determined by the Kibble-Zurek mechanism [110, 193], we find

$$n_M(T) \approx \alpha \left(\frac{T}{T_M}\right)^3 H(T_M)^3, \tag{4.10}$$

where $H(T)$ is the Hubble parameter at temperature T, and $\alpha \gtrsim 1$.[4] The contribution of Eq. (4.9) makes the axion mass effect dominates over the Hubble friction

$$m_a(T) \gtrsim 3H(T), \tag{4.11}$$

below some temperature $T_{\rm i}$ ($\leq T_M$), so that the axion field is subject to damped oscillations for $T \lesssim T_{\rm i}$.

We assume that $U(1)'$ is spontaneously broken at some temperature $T_{\rm f}$ ($\ll T_{\rm i}$), so that monopoles quickly disappear.[5] Axion isocurvature fluctuations are then damped efficiently between temperatures $T_{\rm i}$ and $T_{\rm f}$. Suppose

$$m_a^2(T) \propto T^n, \tag{4.12}$$

($n = 3$ for a constant γ). Since the axion "number density" $m_a(T)\delta\theta(T)^2$ scales as T^3 while Eq. (4.11) is satisfied, we find

$$\delta\theta(T) \propto T^p, \qquad p \approx \frac{6-n}{4}, \tag{4.13}$$

in this period. The final damping factor is thus

$$\Delta \approx \left(\frac{T_{\rm f}}{T_{\rm i}}\right)^{\frac{6-n}{4}}, \tag{4.14}$$

which can be compared with the required amount of damping from observations, Eq. (4.7).

Note that the average axion field $\langle\theta\rangle = \langle a\rangle/f_a$ after the operation of this damping mechanism is determined by the structure of the hidden sector (the original hidden sector $\bar{\theta}$ parameter), which in general differs from the minimum of the late-time axion potential, $\theta_{\rm QCD}$. A homogeneous displacement of the axion field from $\theta_{\rm QCD}$, determining the late-time axion dark matter abundance, is not controlled by the present mechanism, unless we make an extra assumption. For $f_a \gg 10^{12}$ GeV, the value of this displacement must be small, but it can be environmentally selected to be consistent with $\Omega_a \leq \Omega_{\rm DM}$ [120, 177].

[4]Note that α can be much larger than $O(1)$, depending on the dynamics of the phase transition; see e.g. [129]. In this case, monopole-antimonopole annihilations at $T \sim T_M$ may become important; see Section 4.6 for such a scenario.

[5]An alternative possibility will be discussed in Section 4.6.

4.4 Minimal Model

We now consider the minimal model in which the $U(1)'$ sector below T_M contains only a charged scalar field φ, which breaks $U(1)'$ at scale $T_{\rm f}$ ($\ll T_{\rm M}$). In this case, the factor γ in the expression for the induced axion mass, Eq. (4.9), is

$$\gamma \approx \tilde{\gamma}\,\frac{T_M}{f_a}, \tag{4.15}$$

where we have used $T_M \lesssim f_a$, and $\tilde{\gamma} \approx O(1)$ assuming that the $U(1)'$ gauge coupling is of order unity.[6] The axion mass just after the monopole production is then given by

$$\frac{m_a(T_M)}{3H(T_M)} \simeq 0.2\sqrt{\alpha\tilde{\gamma}}\,g_{*M}^{\frac{1}{4}}\sqrt{\frac{T_M^3}{f_a^2\bar{M}_{\rm Pl}}}, \tag{4.16}$$

where we have used $H(T_M) = \rho(T_M)^{1/2}/\sqrt{3}\bar{M}_{\rm Pl}$ and $\rho(T_M) = (\pi^2/30)g_{*M}T_M^4$ with g_{*M} being the effective number of relativistic degrees of freedom at temperature T_M. Assuming that T_M is not much smaller than the unification scale, this number is roughly of order unity (and at least not too much smaller than of order unity). The axion field thus starts having damped oscillations at $T \sim T_i$, within a few orders of magnitude from T_M. Specifically

$$T_i \simeq 1\times 10^{11}\ {\rm GeV}\,\alpha\tilde{\gamma}\sqrt{\frac{g_{*M}}{100}}\left(\frac{10^{16}\ {\rm GeV}}{f_a}\right)^2\left(\frac{T_M}{3\times 10^{15}\ {\rm GeV}}\right)^4. \tag{4.17}$$

Note that if T_i in this expression exceeds T_M, e.g. because of $\alpha \gg 1$, then $T_{\rm i}$ must be set to T_M.

At temperatures below $T_{\rm i}$, axion isocurvature fluctuations are damped. Since Eq. (4.15) implies $n = 3$, so that $p \approx 3/4$ (see Eq. (4.13)),

$$\frac{\delta\theta(T)}{\delta\theta(T_{\rm i})} \approx \left(\frac{T}{T_{\rm i}}\right)^{\frac{3}{4}}. \tag{4.18}$$

Therefore, to avoid the observational constraint of Eq. (4.7), we need

$$T_{\rm f} \lesssim 2\times 10^5\ {\rm GeV}\,\alpha\tilde{\gamma}\sqrt{\frac{g_{*M}}{100}}\left(\frac{\Omega_{\rm DM}}{\Omega_a}\right)^{\frac{2}{3}}\left(\frac{T_M/E_{\rm inf}}{0.3}\right)^4\left(\frac{10^{16}\ {\rm GeV}}{f_a}\right)^{1.5}\left(\frac{E_{\rm inf}}{10^{16}\ {\rm GeV}}\right)^{\frac{4}{3}}, \tag{4.19}$$

where we have used $H_{\rm inf} \approx E_{\rm inf}^2/\sqrt{3}\bar{M}_{\rm Pl}$. We here generate the required value of $T_{\rm f}$ simply by the Brout-Englert-Higgs mechanism associated with φ:

$$V_{\rm hid} = \lambda'\left(|\varphi|^2 - v'^2\right)^2, \tag{4.20}$$

[6] It is important here that the $U(1)'$ sector does not contain a light fermion charged under $U(1)'$. If it did, virtual fermions would partially screen the charge surrounding a monopole, allowing it to spread over a distance or order m_f^{-1}. Here, m_f is the fermion mass. This would suppress the induced mass of the axion so that $\gamma \approx m_f/f_a$ [66]. This will be relevant for models in Section 4.6.

with $v' \approx T_{\rm f}$. We find that unification scale axion dark matter with unification scale inflation can be made consistent by our mechanism.

Incidentally, ignoring $U(1)'$ breaking, we find that monopoles dominate the energy density of the universe at temperature

$$T_* \simeq 6 \times 10^6 \text{ GeV}\, \alpha \sqrt{\frac{g_{*M}}{100}} \left(\frac{T_M}{3 \times 10^{15} \text{ GeV}} \right)^3 \left(\frac{m_M}{3 \times 10^{15} \text{ GeV}} \right), \tag{4.21}$$

which is slightly below the upper bound in Eq. (4.19) in the relevant parameter region. Here, m_M is the monopole mass. This implies that the universe may be monopole dominated toward the end of the damped oscillation period, $T_{\rm f} \lesssim T \lesssim T_{\rm i}$.

4.5 Monopole Annihilations

Here we discuss annihilations of monopoles after $U(1)'$ is spontaneously broken at some temperature T_S ($\sim T_{\rm f}$). After $U(1)'$ is spontaneously broken, monopoles and antimonopoles become connected by strings. For monopole-antimonopole annihilations to occur, the string-monopole system must lose their energies, and there are several processes that can contribute to the energy loss.

We assume the existence of a renormalizable coupling between the $U(1)'$ and standard model sectors, e.g. a quartic coupling between the $U(1)'$ breaking and standard model Higgs fields or a kinetic mixing between $U(1)'$ and $U(1)$ hypercharge:

$$\mathcal{L} \sim \epsilon\, \varphi^\dagger \varphi h^\dagger h, \qquad \epsilon F'_{\mu\nu} F_Y^{\mu\nu}. \tag{4.22}$$

We will find that monopoles quickly disappear, well within a Hubble time, unless the coupling ϵ is significantly suppressed. Note that cosmic strings formed by $U(1)'$ breaking are harmless for $T_S \lesssim 10^{15}$ GeV [162].

Monopole friction

Suppose the correlation length of the $U(1)'$ breaking field, φ, is of order or larger than the average distance between monopoles at $T \sim T_S$:

$$d(T_S) \sim n_M(T_S)^{-\frac{1}{3}} \sim \frac{\bar{M}_{\rm Pl}}{\alpha^{1/3} T_S T_M}. \tag{4.23}$$

In this case, strings will connect monopoles through the shortest possible path, and the energy of a monopole-antimonopole pair to be dissipated is

$$E_0 \sim \eta\, d(T_S) \sim \frac{\bar{M}_{\rm Pl} T_S}{\alpha^{1/3} T_M}, \tag{4.24}$$

where we have estimated the string tension η to be of order T_S^2.

If the monopoles scatter with a thermal bath of temperature T_S through a coupling of strength ϵ, as in Eq. (4.22), then the energy loss rate due to friction is [184]:

$$\dot{E} \sim -\epsilon^2 T_S^2 v^2, \tag{4.25}$$

where v is the velocity of the monopoles, which is given by

$$v \sim \begin{cases} \left(\frac{T_S^2 d(T_S)}{m_M}\right)^{\frac{1}{2}} \sim \left(\frac{T_S \bar{M}_{\rm Pl}}{\alpha^{1/3} T_M^2}\right)^{\frac{1}{2}} & \text{for } T_S \ll \frac{\alpha^{1/3} T_M^2}{\bar{M}_{\rm Pl}}, \\ 1 & \text{for } T_S \gtrsim \frac{\alpha^{1/3} T_M^2}{\bar{M}_{\rm Pl}}, \end{cases} \tag{4.26}$$

where the former and latter cases correspond to nonrelativistic and relativistic monopoles, respectively. In each case, the annihilation timescale $\tau_{\rm ann} \sim |E_0/\dot{E}|$ is given by

$$\tau_{\rm ann} \sim \begin{cases} \frac{T_M}{\epsilon^2 T_S^2} & \text{for } T_S \ll \frac{\alpha^{1/3} T_M^2}{\bar{M}_{\rm Pl}}, \\ \frac{\bar{M}_{\rm Pl}}{\epsilon^2 \alpha^{1/3} T_S T_M} & \text{for } T_S \gtrsim \frac{\alpha^{1/3} T_M^2}{\bar{M}_{\rm Pl}}. \end{cases} \tag{4.27}$$

In both cases, this timescale is of order or shorter than the Hubble timescale, $t_S \sim \bar{M}_{\rm Pl}/T_S^2$, unless ϵ is much smaller than of order unity.

Particle production from strings

If the correlation length of φ is much smaller than the average monopole distance at T_S, then we expect that a string connecting a monopole-antimonopole pair to have a significant number of kinks (from a Brownian formation), and particle production from the string contributes significantly to the dissipation.

Based on the analysis in Ref. [184], we estimate that the power for a string of thickness δ and length L to radiate standard model particles is

$$P \sim \frac{\epsilon^2}{\delta\, \xi(T_S)}, \tag{4.28}$$

per a portion of a string of length $\xi(T_S)$, where $\xi(T_S)$ ($\ll d(T_S)$) is the correlation length of φ.[7] In the case of Brownian strings, the average string length is given by

$$L \sim \frac{d(T_S)^2}{\xi(T_S)}, \tag{4.29}$$

so that the total energy of the string-monopole system to be dissipated and the emission power from it are

$$E_0 \sim \eta L \sim \frac{T_S^2\, d(T_S)^2}{\xi(T_S)}, \tag{4.30}$$

$$\dot{E} \sim P \frac{L}{\xi(T_S)} \sim \frac{\epsilon^2 T_S\, d(T_S)^2}{\xi(T_S)^3}, \tag{4.31}$$

[7] The process of energy dissipation may be much faster, $P \sim \epsilon^2 \eta (\delta/\xi(T_S))^{1/3}$, if cusps form efficiently [37]. Here we adopt a conservative estimate of Eq. (4.28), which is sufficient to eliminate the monopoles quickly.

where we have used $\eta \sim T_S^2$ and $\delta \sim 1/T_S$. The monopole-antimonopole annihilation timescale is thus

$$\tau_{\rm ann} = \frac{E_0}{\dot{E}} \sim \frac{1}{\epsilon^2} T_S\, \xi(T_S)^2 \ll \frac{1}{\epsilon^2} T_S\, d(T_S)^2 \sim \frac{\bar{M}_{\rm Pl}^2}{\epsilon^2 \alpha^{2/3} T_S T_M^2}. \tag{4.32}$$

Again, this is of order or shorter than the Hubble timescale, $t_S \sim \bar{M}_{\rm Pl}/T_S^2$, unless ϵ is much smaller than order unity.[8]

4.6 Technical Naturalness of $U(1)'$

In Section 4.4, we have presented the minimal model in which $U(1)'$ breaking is achieved by a scalar field φ with the potential Eq. (4.20). As it stands, the scale appearing in this potential, v', is not radiatively stable. The radiative stability of this scale is qualitatively and quantitatively different from the problem of protecting the QCD axion from quantum corrections. The $U(1)'$-breaking field φ is a scalar much like the standard model Higgs field whose mass needs to be protected at scales above T_S, unlike the QCD axion whose mass needs to be protected to the level of $\sim 10^{-5} m_a$. Existing ideas to address the hierarchy problem may thus be leveraged to solve this issue. In this section, we discuss extensions/modifications of the minimal model in which the issue of radiative stability does not arise.

Supersymmetric $U(1)'$ sector

One way to construct a technically natural model is to make the $U(1)'$ sector supersymmetric. This requires promoting the $U(1)'$-breaking field φ to chiral superfields $\Phi(+1)$ and $\bar{\Phi}(-1)$. The complication arises because the induced axion mass is suppressed in the presence of light fermions charged under $U(1)'$, as mentioned in footnote 6. To obtain a significant contribution to the axion mass, we need to have a supersymmetric mass for Φ and $\bar{\Phi}$:

$$W = M_\Phi \Phi \bar{\Phi}. \tag{4.33}$$

The breaking of $U(1)'$ is then caused by supersymmetry-breaking squared masses for Φ and $\bar{\Phi}$ of order $\tilde{m}^2 \sim T_S^2$. To maximize the axion mass, we also take $M_\Phi \sim T_S$.[9] The coupling between the $U(1)'$ and the standard model sectors needed for monopole annihilations can be taken as a kinetic mixing between $U(1)'$ and $U(1)$ hypercharge: $\mathcal{L} \sim \epsilon\, [\mathcal{W}'^\alpha \mathcal{W}_{Y\alpha}]_{\theta^2}$ (see Section 4.5). This implies that the standard model is also supersymmetric above the scale $\sim (4\pi/\epsilon)\tilde{m}$.

[8] In the analysis in this subsection, we have ignored the effect of the increase of the relevant correlation length due to interactions of the strings with the thermal bath, which may become important for $T_S \lesssim T_M^2/\bar{M}_{\rm Pl}$. In this case, however, the analysis in the previous subsection applies, which also says that monopoles quickly disappear after $U(1)'$ symmetry breaking.

[9] The coincidence of the scales $\tilde{m}$ and M_Φ is analogous to the μ problem in the minimal supersymmetric standard model, which can be addressed, e.g., as in Ref. [71, 42].

With this setup, the induced axion mass is given by Eq. (4.9) with

$$\gamma \approx \frac{M_\Phi}{f_a} \sim \frac{T_S}{f_a}. \tag{4.34}$$

Plugging this into Eq. (4.19) with $T_S \sim T_{\rm f}$, we find that α must be much larger than 1 for the model to work. We thus suppose that the dynamics of the phase transition producing monopoles is such that $\alpha \gg 1$. The largest possible abundance of monopoles obtained in this case is determined by the freezeout abundance (instead of Eq. (4.10)), which is given by [153]

$$n_M(T) \approx \left(\frac{T}{T_M}\right)^3 \frac{\sqrt{g_{*M}} T_M^4}{\bar{M}_{\rm Pl}}, \tag{4.35}$$

where we have assumed an $O(1)$ $U(1)'$ gauge coupling. The axion mass at $T \sim T_M$ is then

$$\frac{m_a(T_M)}{3H(T_M)} \sim \frac{\sqrt{T_S \bar{M}_{\rm Pl}}}{g_{*M}^{1/4} f_a}, \tag{4.36}$$

so that the axion field starts damped oscillations at

$$T_{\rm i} \sim \frac{T_S T_M \bar{M}_{\rm Pl}}{\sqrt{g_{*M}} f_a^2}. \tag{4.37}$$

This gives the damping factor of

$$\Delta \approx \left(\frac{T_{\rm f}}{T_{\rm i}}\right)^{\frac{3}{4}} \sim \left(\frac{f_a^2}{T_M \bar{M}_{\rm Pl}}\right)^{\frac{3}{4}}. \tag{4.38}$$

We find that the mechanism is not as strong as in the minimal model, but it can still save the scenario with f_a, T_M, $E_{\rm inf}$ as large as $\sim 10^{15}$ GeV.

Possibility of unbroken $U(1)'$

We finally mention an alternative (and very different) possibility that $U(1)'$ monopoles may be efficiently eliminated without breaking $U(1)'$. This may happen if the monopole under consideration is in fact a dyon that also carries a charge under a hidden non-Abelian gauge group G' (to which the axion field does not couple). In this case, if G' confines at a scale Λ', then dyons can be subjected to extra strong annihilation processes.

Suppose the G' sector contains light particles that are electrically charged under G'. When G' confines at $T \sim \Lambda'$, dyons pick up these light particles, becoming G' hadrons. At this point, the dyon-antidyon annihilation cross section is expected to become large $\sim 1/\Lambda'^2$, as in the analogous situation for a heavy stable colored particle [107]. This will efficiently eliminate dyons if the confinement scale is sufficiently low $\Lambda' \lesssim 100$ TeV, giving $T_{\rm f} \sim \Lambda'$. Since this scenario does not require breaking of the $U(1)'$ symmetry, the $U(1)'$ sector need not have a light charged scalar or fermion, which would, respectively, lead to the issue of radiative stability and axion mass suppression. Further studies of this possibility, including a detailed analysis of whether dyon annihilation is indeed strong enough, are warranted.

4.7 Conclusions

In this paper we have presented a mechanism that suppresses axion isocurvature fluctuations due to the dynamics of a hidden $U(1)'$ sector coupled to the axion field. In particular, this sector produces $U(1)'$ monopoles at $T \sim T_M$, which disappear at $T \sim T_{\rm f}$ ($\ll T_M$). For temperatures between $T_{\rm i}$ ($\gg T_{\rm f}$) and $T_{\rm f}$, the effective axion mass induced by the monopoles makes the axion heavier than the Hubble parameter, so that the isocurvature fluctuations are damped. Since the average value of the axion field after the damping is not necessarily at the minimum of the zero-temperature potential determined by QCD, homogeneous coherent oscillations after the QCD phase transition may still produce axion dark matter [120, 177].

We have presented a minimal model in which this mechanism successfully operates. This model accommodates a large enough time interval in which the axion isocurvature fluctuations are damped, so that axion dark matter with a unification scale decay constant can be consistent with unification scale inflation. We have also discussed extensions/modifications of the minimal model in which the issue of radiative stability does not arise.

Since the axion provides a leading solution to the strong CP problem, it is important to fully study its consistency. If a future CMB experiment discovers inflationary gravitational wave signals, it would exclude naive axion models with the Peccei-Quinn symmetry broken before the end of inflation. Our mechanism makes the QCD axion alive even in such a case, without requiring the Peccei-Quinn symmetry breaking scale to be below the inflationary scale. This is particularly important for a string axion, which has a virtue that explicit breaking of the Peccei-Quinn symmetry (which needs to be extremely small to solve the strong CP problem [106]) is generated only at a nonperturbative level [171, 11]. Our mechanism allows for a string axion to be a consistent solution to the strong CP problem even if inflationary gravitational wave signals are discovered, and it would also keep open the possibility that axion dark matter may be discovered by high precision experiments such as those proposed in Ref. [72, 73, 40].

Chapter 5

A Holographic Entanglement Entropy Conjecture for General Spacetimes

5.1 Introduction

A theory of quantum gravity should not apply only to asymptotically locally anti-de Sitter (AlAdS) spacetimes. For this reason, the AdS/CFT correspondence [122, 189], although immensely successful, has fallen short of a description of the quantum mechanics of spacetime. The AdS restriction is severe: Maldacena's conjecture does not apply in an obvious way to even the cosmological spacetime we find ourselves in.

If a quantum theory applies to general spacetimes, it is desirable that it reduces to AdS/CFT in the appropriate cases. This suggests a strategy for guessing properties of a complete theory: consider specific aspects of AdS/CFT and devise generalizations that are applicable to other spacetimes. If one knew only of Special Relativity, she could guess aspects of General Relativity by thinking to "promote" the flat metric to a dynamical one. Similar statements can be made about the relation between many other pairs of theories. But retrospective examples obscure the challenge: one cannot confidently know what to promote (and how to promote it) to enlarge the regime of validity of a given theory.

Holographic entanglement entropy, proposed by Ryu and Takayanagi (RT) [157], proved by Lewkowycz and Maldacena [113], and made covariant by Hubeny, Rangamani, and Takayanagi (HRT) [99], is a beautiful property (or, in the covariant case, conjecture) of AdS/CFT. Below we describe a promotion of holographic entanglement entropy beyond the scope of AdS/CFT that applies just as well to cosmological spacetimes as it does to asymptotically AdS spacetimes. In the case of the latter, it reduces to the HRT proposal. Moreover, the promoted holographic entanglement entropy satisfies, for nontrivial reasons, expected properties of entanglement entropy like strong subadditivity.

The HRT prescription provides a way to compute entanglement entropy of a spatial region A in a quantum state dual to an AlAdS spacetime. The procedure is to consider ∂A, the boundary of the spatial region, and to find the area of a codimension 2 extremal surface

that is anchored to ∂A. A naïve extension of this idea to general spacetimes would be to take A to be a region in the conformal boundary of an arbitrary spacetime. This approach fails: what is the boundary of a closed FRW universe with past and future singularities?

In our proposal, we anchor extremal surfaces to a *holographic screen*. Holographic screens are codimension 1 surfaces that appear to be the most natural place for quantum states dual to arbitrary geometries to live on. In fact, they were proposed by Bousso [28] in an attempt to find the analogue of the AdS boundary when extending holography to general spacetimes. If one believes the covariant entropy bound [27], then there is essentially no other reasonable class of surfaces for this purpose.

Outline. In section 5.2 we first review the concept of holographic screens [28] with an emphasis on the recent developments of Bousso and Engelhardt [31, 32] which identified a class of screens that satisfy an area monotonicity law. We then give the definition of *holographic screen entanglement entropy* and list a number of its key properties. We conclude the section by stating our *screen entanglement conjecture*—a proposal that holographic screen entanglement entropy actually measures von Neumann entropy in a putative holographic description of general spacetimes. Section 5.3 contains technical developments including proofs of the properties of screen entanglement entropy that are advertised in section 5.2. Section 5.4 gives cosmological examples of holographic screens and their extremal surfaces. We focus particularly on FRW universes that approach de Sitter space at late times. Section 5.5 concludes by reviewing the procedure for computing screen entanglement entropy and by suggesting extensions to our proposal such as possible methods for computing subleading contributions to holographic screen entanglement entropy.

5.2 Holographic Screen Entanglement Entropy

We open this section with a brief review of holographic screens, especially past and future holographic screens. Readers that are already familiar with the content [27, 28, 31, 32] may still find it useful to read through these paragraphs to become familiar with our conventions and notation. Throughout this paper we will work in a globally hyperbolic spacetime M of dimension d that satisfies the null energy condition. We assume that the spacetime satisfies the genericity conditions laid out in [28, 32].

Suppose that B is an orientable spacelike codimension 2 submanifold of M. It is possible to find an independent pair of future directed null vector fields on B that are everywhere orthogonal to B. If one of these vector fields has vanishing null expansion on B, we will say that B is *marginal*. If one vector field has zero expansion on B while the other has negative (positive) expansion on B, we say that B is *marginally trapped (marginally anti-trapped)*.

A *past holographic screen* is a codimension-1 submanifold $\mathcal{H}$ of the spacetime that is foliated by marginally anti-trapped compact spacelike surfaces called *leaves*. The foliation into leaves is unique: other splittings of $\mathcal{H}$ cannot satisfy the marginally anti-trapped condition.

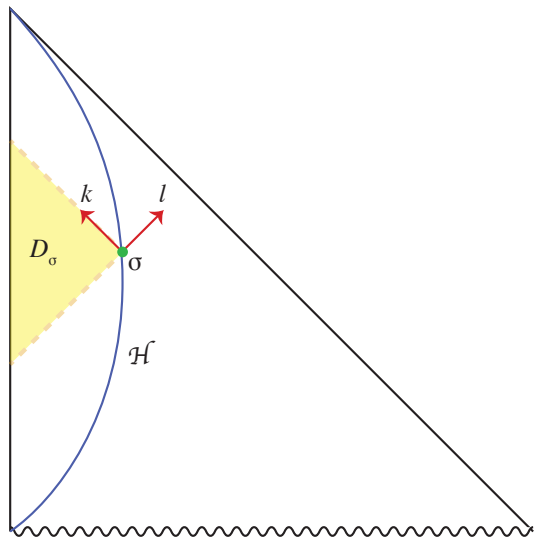


Figure 5.1: An example of a past holographic screen $\mathcal{H}$. One particular leaf σ is highlighted here along with its null orthogonal vector fields k and l satisfying $\theta^k = 0$ and $\theta^l > 0$. The causal region D_σ plays a critical role in our generalization of holographic entanglement entropy.

A *future holographic screen* is instead foliated by marginally trapped surfaces. In this paper, we will always assume that leaves have the topology of S^{d-2}.

Holographic screens are generated by null foliations: if $\{N_r\}$ is a null foliation of a spacetime, it is possible to identify a family of leaves $\{\sigma(r)\}$ with $\sigma(r) \subset N_r$ by finding the codimension 2 surface of maximal area on each null surface. In general, this will break the values of the parameter r into open intervals, some of which correspond to past holographic screens and others corresponding to future screens.[1] Isolated values of r that lie between past and future screens correspond to the case where $\sigma(r)$ is an isolated extremal sphere which can join a past and future screen. Such a sphere will not be considered to lie on a past or future holographic screen by convention. This occurs in the case of a closed universe with a big crunch: see figure 5.9.

Some of the simplest examples of holographic screens arise in the "observer-centered" case where we take $\{N_r\}$ to be the set of past light-cones of an observer's worldline in some spacetime. In the case of FRW cosmology with the observer taken to be comoving, such holographic screens are just apparent horizons. Figure 5.1 shows an example of such a holographic screen. See also figures 5.6 (top) and 5.9. Because past holographic screens are often generated in this way, we will mostly focus on the case of past screens throughout

[1] It is also possible that for some values of r, $\sigma(r)$ does not have a definite sign for θ^l. We leave the investigation of this scenario to future work.

this paper. However, all results below apply equally well to future screens with appropriate modifications.

Because null foliations are highly non-unique, holographic screens are also non-unique. For example, in the observer-centered case, a past holographic screen can be obtained obtained by considering the surfaces of maximal area on the past light-cones of an observer's worldline if the maximal area surfaces are anti-trapped and compact which we assume. In this case, performing a modification to the worldline will modify the holographic screen.[2] From this point of view, holographic screens appear to be "pro-complementarity" objects. The potential importance of this aspect of screens is further discussed below.

Suppose that $\mathcal{H}$ is a past holographic screen. Let σ be a leaf of $\mathcal{H}$ and let k and l denote, respectively, the ingoing and outgoing future-directed null surface-orthogonal vector fields on σ. (It may be useful to refer to figure 5.1.) Then, the condition that σ is marginally anti-trapped means that

$$\begin{aligned} \theta^k &= 0 \\ \theta^l &> 0 \end{aligned} \tag{5.1}$$

where θ^k and θ^l denote the expansion of congruences in the k and l directions at σ.

Every holographic screen comes with a *fibration*. A fibration is a family of curves generated by a nonvanishing vector field h on and tangent to $\mathcal{H}$ with the property that h is orthogonal to every leaf. If we extend the vector fields k and l to all of $\mathcal{H}$ (so that they are surface orthogonal to every leaf), then $h = \alpha l + \beta k$ where α and β are scalar functions on $\mathcal{H}$. h is not required to be timelike, spacelike or null and, in fact, can switch between these three cases on one screen. Thus, holographic screens need not have definite signature.[3] Lacking a definite signature, normalization of h is arbitrary. Nonetheless, it is convenient to write the leaves of $\mathcal{H}$ as $\sigma(r)$ where r is some (non-unique) parameter and to then normalize h by the condition $dr(h) = 1$.

Bousso and Engelhardt proved that $\alpha > 0$ at every point in $\mathcal{H}$ and concluded that leaves have strictly increasing area [31, 32]. More precisely, the area of $\sigma(r_2)$ is greater than the area of $\sigma(r_1)$ if $r_2 > r_1$. In fact, if $\|\cdot\|$ denotes the area functional, then

$$\frac{d}{dr}\|\sigma(r)\| = \int_{\sigma(r)} d^{d-2}y\sqrt{g^{(\sigma(r))}}\,\alpha\,\theta^l$$

[2]Note that the non-uniqueness of holographic screens for a given spacetime fits well with the ideas of [140, 137, 136] where a strong emphasis is placed on the importance of "fixing the gauge" in quantum gravity. This was clearly discussed in [140] in which the role of a gauge-fixed apparent horizon (essentially a holographic screen though not a past or future screen) was discussed. We do not commit to the pictures described in these papers.

[3]This is the key distinguishing feature between past (and future) holographic screens and related objects including future outer trapping horizons and dynamical horizons [89, 90, 13, 12] that were introduced in an attempt to find a "quasi-local" definition of a black hole. Past and future holographic screens can be regarded as a synthesis such ideas with those of [27].

which is positive by equation 5.1 and the fact that $\alpha > 0$. Here, $g^{(\sigma(r))}$ denotes the induced metric on $\sigma(r)$. Note, in particular, that the area is *strictly* increasing for all intervals of r. The inequality would not be strict if it were not for the genericity conditions of [32].

Definition and Properties of Holographic Screen Entanglement Entropy

As before, let $\mathcal{H}$ denote a past holographic screen. Everything below can be modified to the case of a future holographic screen without difficulty.

It is helpful to emphasize the following result which follows from the genericity conditions of [32]:

- *Strict Focusing.* If B is a codimension 2 spacelike surface, the four surface-orthogonal null congruences have strictly decreasing expansion as they move away from B.

This means that there is always enough matter content everywhere in the spacetime to focus neighboring null geodesics. If M fails to satisfy this condition, it can be made to do so by sprinkling a very small amount of classical matter everywhere.

As discussed above, there is a unique foliation of $\mathcal{H}$ into anti-trapped leaves. Let σ be a particular leaf in this foliation and let k and l denote the vector fields on σ that satisfy equation 5.1. Because M is globally hyperbolic, there exists a Cauchy surface S_0 containing σ such that $S_0 \backslash \sigma$ consists of a disconnected interior and exterior. The interior of S_0 is defined so that a vector on σ pointing toward the interior takes the form $c_1 k - c_2 l$ with $c_1, c_2 > 0$. Let S denote the union of the interior of S_0 with σ. We will assume that S is compact and that it has the topology of a solid ball. Now let D_σ be the domain of dependence of S, $D_\sigma = D(S)$, with the convention that D_σ includes orthogonal null surfaces generated by k and $-l$.

Suppose that A is a $d-2$ dimensional submanifold of σ with a boundary. Consider the set of extremal codimension 2 surfaces that are anchored to and terminating at ∂A, and contained entirely in D_σ (see figure 5.2). In section 5.3 we will give conditions on D_σ that ensure that this set is not empty. Taking the existence of such a surface for granted, let the one of minimal area be denoted by ext (A) and define the *holographic screen entanglement entropy* (or *screen entanglement entropy* for brevity) of A as

$$S(A) = \frac{\|\text{ext}\ (A)\|}{4}. \tag{5.2}$$

The quantity $S(A)$ is the most natural generalization of the HRT proposal to general spacetimes. We emphasize that we have defined screen entanglement entropy geometrically without reference to a quantum theory. The term "entanglement entropy" is only meant suggestively. Nonetheless, below we state a *screen entanglement conjecture*: that $S(A)$ is in fact the von Neumann entropy of a subsystem of a holographic quantum state for general spacetimes. Regardless of the validity of this conjecture, we are free to study $S(A)$ as we

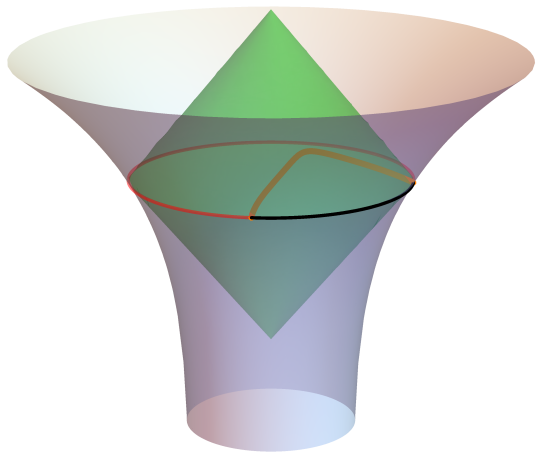


Figure 5.2: This figure depicts our construction of holographic entanglement entropy in general spacetimes. The horn-shaped surface is a past holographic screen $\mathcal{H}$. The black and red codimension 2 regions together form a single leaf σ. The black segment represents a region A and the extremal surface ext (A) (orange) is anchored to its boundary. The causal region D_σ is the green diamond (both interior and boundary). Note that ext $A \subset D_\sigma$.

have defined it. As we will see, the properties of holographic screens ensure that screen entanglement entropy possesses numerous properties reminiscent of von Neumann entropy which we now discuss.

Properties of Holographic Screen Entanglement Entropy and Extremal Surfaces

- *Existence and Containment.* In section 5.3 we provide conditions for ext (A) to exist. This is a nontrivial issue because of the "containment condition" that ext $A \subset D_\sigma$. Arguments that D_σ contains an extremal surface rely critically on the assumption that A is in a leaf of a holographic screen. Moreover, the condition that ext $(A) \subset D_\sigma$ gives rise to properties of holographic screen entanglement entropy like strong subadditivity (see below) and will allow us to reasonably define an entanglement wedge for A. For an example of the importance of the containment condition, see equation 5.4 below and the paragraphs around it.

- *(Strong) Subadditivity.* Suppose that A and B are regions in σ. Then,

$$S(A) + S(B) \geq S(A \cup B) + S(A \cap B)$$

  where S is the function defined in 5.2. This result holds regardless of whether or not A and B intersect as long as we take the convention that $S(\emptyset) = 0$. As we will see in section 5.3, the proof of this is a modified version of Wall's [187] "maximim" proof for the HRT case. This does not mean that strong subadditivity is an obvious result: most of the work in section 5.3 is to show that the properties of leaves of holographic screens are sufficient to generalize Wall's arguments to our context.

- *Page Bounded.* Define the extensive entropy of A as $S_{\text{extensive}}(A) = \|A\|/4$. Then, the holographic screen entanglement entropy satisfies the following *Page bound*:[4]

$$S(A) \leq \min\{S_{\text{extensive}}(A), S_{\text{extensive}}(\sigma \setminus A)\}. \tag{5.3}$$

  This is a simple consequence of the maximin construction we give in section 5.3. Note that the area law for holographic screens implies that this inequality becomes a weaker constraint if we transport A along the fibration vector field defined above. In certain cases, the inequality saturates and $S(A)$ approaches a "random entanglement limit." (See section 5.4 for examples of this in cosmology.)

- *Reduction to the HRT Proposal.* As explained in detail in [28], the AdS boundary can be regarded as a holographic screen. In this case, surfaces of constant time in the dual field theory correspond to leaves, and our proposal becomes identical to the covariant holographic entanglement entropy conjecture of [99].

The Screen Entanglement Conjecture

We are now in a position to state our conjecture about the role of $S(A)$ in quantum gravity. This conjecture is the primary concern of this paper. Nonetheless, we emphasize that the mathematical developments below (e.g. the proof that $S(A)$ satisfies standard properties of von Neumann entropy) do not rely on any conjectural statements.

Our proposal can be regarded as an extension of a covariant holographic principle due to Bousso which we now review very briefly. In [28], Bousso integrated the ideas of [173, 169] with his covariant entropy conjecture [27] and proposed that each marginal surface B foliating a holographic screen is associated with a Hilbert space $\mathcal{H}_B$ of dimension $\exp(\text{area}(B)/4)$ and that states in $\mathcal{H}_B$ holographically define the state on a null surface N passing through B in the marginal direction. For our purposes, this holographic principle takes the following form. To each leaf σ of a past or future holographic screen we assign a density matrix ρ_σ. The density matrix acts on a Hilbert space of dimension $\exp(\text{area}(\sigma)/4)$ which may be a subspace

[4] The term "Page bound" is motivated by Page's considerations of the entanglement entropies of subsystems [144].

of a "complete" Hilbert space.[5] The covariant entropy bound suggests that ρ_σ encodes the quantum information on the null slice generated by k and $-k$ where k is the null vector field with $\theta^k = 0$ on σ.

We now assume Bousso's holographic principle and state our new conjecture. We propose that every region A of σ (up to string scale resolution) corresponds to a subsystem of the Hilbert space that ρ_σ acts on. We conjecture that the von Neumann entropy of that subsystem in the density matrix ρ_σ is given, at leading order, by $S(A)$ as we have defined it in equation 5.2.

We refer to this statement as the *screen entanglement conjecture*. Because a holographic quantum theory dual to arbitrary spacetimes is not known, the screen entanglement conjecture is not a mathematical statement about the relation between two known theories (as in the case of HRT). Instead, our conjecture suggests a way to compute properties of quantum states in an unknown theory. It is our hope that this will, in fact, be a step toward developing a quantum theory for arbitrary spacetimes.

Nonuniqueness of Holographic Screens and Frame-Dependence in Quantum Gravity

It was emphasized above that in a given spacetime, there is no unique preferred holographic screen. As a consequence, screen entanglement entropy cannot even be defined before first deciding on a particular choice of a screen. This might seem to put the screen entanglement conjecture on haphazard footing, but we explain here why this arbitrariness is, in fact, a necessary feature of any generalization of holographic entanglement entropy to general spacetimes.

Conventional holographic entanglement entropy in AdS/CFT is reference frame dependent in the following sense. Consider an observer in an asymptotically AdS spacetime M with conformal boundary ∂M following a worldline $p(\tau)$. Here, τ is the proper time parameter of the observer. At a given value of τ, we can consider a spacelike *cut* of the boundary [130, 61]:

$$C(\tau) = \partial J_-\big(p(\tau)\big) \cap \partial M.$$

Here, $J_-(q)$ denotes the causal past of a point q. A region A_Ω on $C(\tau)$ can be specified by considering a portion Ω of a small sphere on the tip of the past light cone of the point $p(\tau)$ and following points in Ω down null geodesics until ∂M is reached. Thus, once the trajectory $p(\tau)$ is decided upon, we can use the HRT formula to compute $S(A_\Omega, \tau)$, the holographic entanglement entropy of the region A_Ω on the cut $C(\tau)$. If the trajectory is changed, $S(A_\Omega, \tau)$ correspondingly transforms. At the level of the dual CFT, this discussion corresponds to the

[5] The concept that the states corresponding to any particular approximately fixed geometry form a subspace of a complete Hilbert space is due to Nomura [132, 133]. In his formulation, a larger Hilbert space for arbitrary geometries is a direct sum over subspaces for each geometry. This direct sum itself is only a subspace of the complete Hilbert space which may include an "intrinsically stringy" subspace with no geometrical interpretation. This construction may provide insight into how quantum mechanics can be unitary despite the fact that screens have non-constant area.

fact that quantum states and their time-dependence have a gauge-redundancy that is fixed by making a choice of time-slicing on the boundary ∂M.

In the case of the screen entanglement conjecture and a spacetime that is not asymptotically AdS, a null foliation must be selected to fix a holographic screen. As discussed above, a simple way to do this is to choose a curve $p(\tau)$, and, at any given τ, follow along the past light-cone of $p(\tau)$ until a marginal surface $\sigma(\tau)$ is obtained. The role of $\sigma(\tau)$ is analogous to that of the cut $C(\tau)$ in asymptotically AdS spacetimes. The foliation dependence of screen entanglement entropy is closely tied to the frame-dependence of the HRT formula. This is an example of "fixing the gauge" in quantum gravity, a concept developed in [133].

In the case of asymptotically AdS spacetimes, no matter what worldline $p(\tau)$ is chosen, the union of all of its cuts will always be a subset of the boundary.[6] In general spacetimes however, the particular holographic screen obtained by taking the union over all τ of $\sigma(\tau)$ will depend on the choice of the worldline. Thus, the surface on which holographic quantum states are defined is no longer tethered to the spacetime. This is a basic property of Bousso's holographic principle, not one that arises only in the more extended framework of this paper. We regard this aspect of holographic screens as being in the spirit of black hole complementarity, where quantum information is not attached to a fixed spacetime position (e.g. a qubit is not inside or outside a black hole) until an observer is selected to describe the system.

5.3 Proofs of Strong Subadditivity and Other Relations

In this section we prove key technical results about holographic screen entanglement entropy including many of the properties advertised above. The notation and conventions we will use are the same as those given in section 5.2. In particular, $\mathcal{H}$ is a past holographic screen in a globally hyperbolic spacetime of dimension d that satisfies the genericity conditions of [32]. σ is a compact leaf of $\mathcal{H}$ which we assume to have the topology of S^{d-2}. k and l are null orthogonal vector fields on σ satisfying equation 5.1. S_0 is a Cauchy slice containing σ and S is the portion of S_0 that is enclosed by σ including σ itself (the enclosed side is defined in section 5.2). S is assumed to have the topology of a compact $d-1$ ball. D_σ is the domain of dependence of S.

As always, the case of a future holographic screen is omitted because it presents no additional subtlety.

[6] This follows trivially from the definition of $C(\tau)$. However, in some cases past directed null geodesics of $p(\tau)$ may fail to reach ∂M.

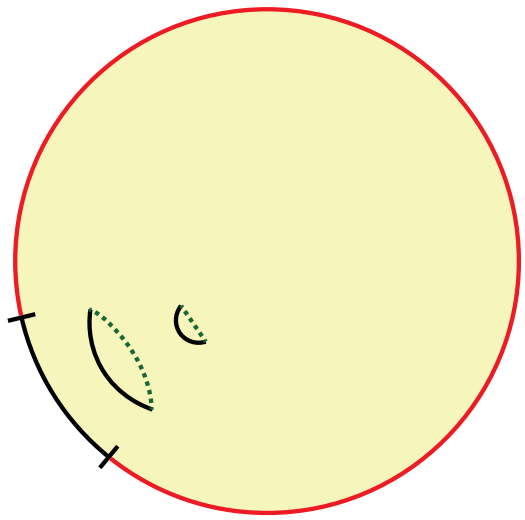


Figure 5.3: The proof of lemma 1 involves a continuous family of surfaces A_s along with their extremal surfaces (dotted curves).

Existence and Containment of Extremal Surfaces

As discussed in section 5.2, it is nontrivial and critical to show the existence of an extremal surface anchored to ∂A that lies entirely in D_σ. We now prove that such a surface exists under very generic conditions. Our first step is to show that ext (A) exists in the case that D_σ is compact. This is a common situation[7] although it is not the case if the ingoing light sheets of σ encounter a singularity.

Lemma 1. *If D_σ is compact, then there exists a codimenson 2 extremal surface anchored and terminating at ∂A that lies entirely in D_σ and that intersects ∂D_σ only at ∂A.*

Proof. Let Σ_+ and Σ_- denote the future and past ingoing light-sheets of σ. We now extend Σ_- to a slightly larger light-sheet, $\tilde{\Sigma}_-$, by following the future directed null congruence of l. Because $\theta^l > 0$ on σ, we can make this extension so that $\tilde{\Sigma}_-$ has $\theta^l > 0$ everywhere and so that there exists an open set in $\tilde{\Sigma}_-$ containing σ.

In the language of [62], both $\Sigma_+ \backslash \sigma$ and $\tilde{\Sigma}_-$ are extremal surface barriers because they have negative expansion in the k and $-l$ directions respectively. Moreover, $\partial D_\sigma \subset (\Sigma_+ \setminus \sigma) \cup \tilde{\Sigma}_-$.

[7] Suppose that the future and past ingoing light-sheets of σ terminate at caustics rather than singularities. Let C_+ and C_- denote the set of the first caustics encountered (local or nonlocal) by null geodesics in the future and past light sheets respectively. Then, if $D_\sigma = J_-(C_+) \cap J_+(C_-)$, we can conclude that D_σ is compact. This follows from the fact that $C_\pm$ inherits the compactness of σ and from the fact that global hyperbolicity implies that $J_-(K_1) \cap J_+(K_2)$ is compact if K_1 and K_2 are compact.

It follows that ∂D_σ is itself an extremal surface barrier for extremal surfaces in the interior[8] of D_σ.

Now consider the region A. The spherical topology[9] of σ ensures that it is possible to introduce a continuous one-parameter family of submanifolds of D_σ, A_s, such that

- A_0 consists of a single point in the interior of D_σ
- $A_1 = A$
- for $0 < s < 1$, A_s is a codimension 2 submanifold of the interior of D_σ that is diffeomorphic to A.

This is shown in figure 5.3. Note, in particular, that if $s < 1$, $A_s \cap \partial D_\sigma = \emptyset$.

If $\epsilon > 0$ is sufficiently small, then the extremal surface of minimal area that is anchored to ∂A_ϵ lies entirely in the interior of D_σ. Denote this extremal surface by $\Gamma(\epsilon)$. Consider increasing the value of the parameter s from ϵ to 1. For each value of s, construct an extremal surface $\Gamma(s)$ (not necessarily the one of minimal area) anchored to ∂A_s. The compactness of D_σ (which ensures that it is bounded and has no singularities) together with the fact that, as discussed above, ∂D_σ is an extremal surface barrier, allows us to take $\Gamma(s)$ to not jump discontinuously and to be contained in the interior of D_σ for all $s < 1$. When we take the limit sending s to 1, the extremal surface anchored to ∂A must intersect ∂D_σ at ∂A and nowhere else: if it did intersect ∂D_σ outside of ∂A, the extremal surface would be locally tangent to an extremal surface barrier with strictly nonzero null extrinsic curvature.

□

The unwanted assumption that D_σ is compact (which fails in the event that Σ_+ or Σ_- encounter a singularity) can be dropped if there exists a codimension 0 submanifold (with boundary) of D_σ, R, which "restricts" extremal surfaces (see figure 5.4). By this we mean that

1. R is compact,
2. There exists an open set U containing S with $D_\sigma \cap U = R \cap U$, and
3. $\partial R = (\partial D_\sigma \cap R) \cup B$ where B is an extremal surface barrier for codimension 2 extremal surfaces inside in R.

These conditions for R are designed to ensure that R can be used in lemma 1 in place of D_σ without difficulty. The existence of such regions R relies on the existence of the barrier

[8] In [62], extremal surfaces are confined to regions referred to as the "exterior" of an extremal surface barrier. The interior of D_σ, i.e. $D_\sigma \setminus \partial D_\sigma$, is analogous to exterior regions studied by Wall and Engelhardt.

[9] We remind the reader that our conventions are those laid out in the first paragraph of section 5.2. In particular, we are making simplifying topological assumptions about σ and S. We will leave it to future work to investigate the consequences of relaxing these assumptions.

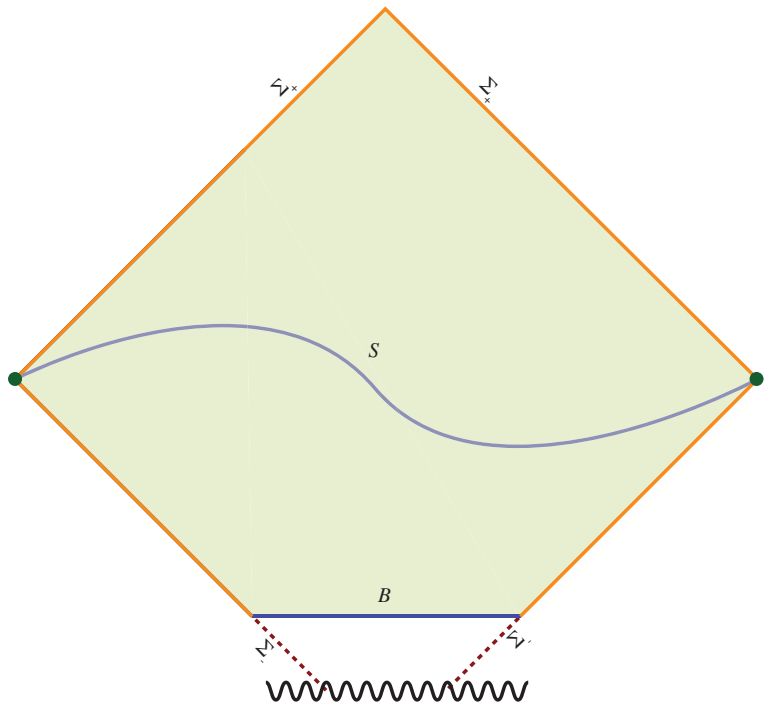


Figure 5.4: The idea of a compact restriction is shown here. The restriction R is the shaded region along with its boundary, the blue and orange lines. ∂R consists of two parts: an extremal surface barrier B (blue) and a portion of ∂D_σ (orange). In this figure, the barrier B protects extremal surfaces in R from a singularity. Not shown are extremal surfaces in R, none of which contact ∂R except at their anchor on the leaf σ.

B. The arguments in theorem 11 of [187] show that Kasner singularities are always protected by such barriers. Hartman and Maldacena [83] encountered a barrier protecting black hole singularities from codimension 2 extremal surfaces. Constant time slices in FRW spacetimes are another example of suitable barriers.[10]

Any region $R \subset D_\sigma$ satisfying the conditions will be called a *compact restriction* of D_σ. Note that, in particular, if D_σ is compact then D_σ is a compact restriction of itself. Our findings can now be summarized by the following improvement upon lemma 1:

Theorem 1. *If D_σ possesses a compact restriction, then there exists a codimenson 2 extremal surface anchored and terminating at ∂A that lies entirely in D_σ and that intersects ∂D_σ only at ∂A.*

[10]Many extremal surfaces are anchored at singularities and thus pass through barriers. This is irrelevant because the barriers we are discussing here play the of ∂D_σ in the proof of lemma 1. As a region A_s is deformed from a point inside R into $A \subset \sigma$, extremal surfaces anchored to ∂A_s cannot smoothly pass B or ∂D_σ.

To better appreciate this theorem, it is helpful show that the statement is false if σ is not a leaf of a holographic screen. Consider $2+1$ dimensional Minkowski space with inertial coordinates (t, x, y) and let $\mathcal{C}$ denote the large cylinder satisfying $x^2 + y^2 = R^2$ with $R \gg 1$. Consider the two line segments on $\mathcal{C}$ that are approximately given by

$$\begin{aligned} A &= \{(t = \frac{1}{2}|x|, -1 > x \geq 0, y = R)\} \\ B &= \{(t = \frac{1}{2}|x|, 0 \leq x < 1, y = R)\} \end{aligned} \tag{5.4}$$

and construct any spacelike "time slice" on $\mathcal{C}$, σ, that includes AB. It is easy to see that the extremal surface anchored to $\partial(AB)$ is a straight line that is timelike related to AB and thus fails to lie within the domain of dependence D_σ. To see how severe this problem is, note that the segments A and B fail to satisfy subadditivity of entanglement entropy. That is, the inequality $S_A + S_B \geq S_{AB}$ is false. Note that in this example σ fails to satisfy equation 5.1 because of the kink at $A \cap B$.

A Maximin Construction for Holographic Screens

Theorem 1 ensures that holographic screen entanglement entropy is a well-defined quantity in a broad set of cases. We will now demonstrate that this quantity satisfies expected properties of entanglement entropy. To do this, it is very useful to closely follow [187] and introduce a maximin construction of ext A. Our construction will be slightly modified from that used for HRT surfaces anchored to the AdS boundary. Wall's maximin prescription involves considering a collection of Cauchy slices that are anchored only to ∂A. Because we already know that ext A lies inside of D_σ, we will introduce a stronger constraint requiring that we only consider achronal slices that are anchored to all of σ.

Definition and Existence of Mm(A)

Our setup remains unchanged. Fix a past (or future) holographic screen $\mathcal{H}$ in a globally hyperbolic spacetime and let σ be a leaf. We take a Cauchy surface S_0 containing σ and define S as the closure of the portion of S_0 inside of σ. As before, we require that S is compact and that it has the topology of a solid $d-1$ ball. Let $D_\sigma = D(S)$. We also fix a region A in σ with a boundary. Now define $\mathcal{C}_\sigma$ as the collection of codimension 1 compact achronal surfaces that are anchored to σ and that have domain of dependence D_σ. Note, in particular, that $S \in \mathcal{C}_\sigma$. Moreover, note that the global hyperbolicity of D_σ ensures that every element of $\mathcal{C}_\sigma$ has the same topology as S: that of a compact $d-1$ ball.

Take any $\Sigma \in \mathcal{C}_\sigma$. Let $\min(\partial A, \Sigma)$ denote the codimension 2 surface of minimal area[11] on Σ that is anchored to ∂A. The existence of $\min(\partial A, \Sigma)$ is guaranteed by the compactness of Σ

[11] Wall [187] added the condition that $\min(\partial A, \Sigma)$ be homologous to A. While this condition ought to be included in our discussion as well, the assumption that S (and thus every element of $\mathcal{C}_\sigma$) has the topology of a compact $d-1$ ball makes a homology condition trivial. We leave the task of investigating more general topologies to future work.

and theorem 9 of [187]. Define a function $F : \mathcal{C}_\sigma \to [0, \text{area}(A)]$ by $F(\Sigma) = \text{area}(\min(\partial A, \Sigma))$. Now assume that there exists a Σ_0 in $\mathcal{C}_\sigma$ that maximizes F (globally). We now define $\min(\partial A, \Sigma_0)$ as the maximin surface of A, and we will denote it by $\text{Mm}(A)$. If there are several maximin surfaces, $\text{Mm}(A)$ can refer to any of them.

The existence of $\text{Mm}(A)$ can be proven in many cases by appropriately importing the arguments of theorems 10 and 11 in [187] which we only briefly describe here. Consider the Cauchy surface S_0 which can be identified as a slice in a foliation of Cauchy surfaces $\{S_t\}$. Using this definition of time, we can identify a surface $\Sigma \in \mathcal{C}_\sigma$ with a function $t_\Sigma : S_0 \to \mathbf{R}$ in a natural way: if I_x denotes the integral curve of ∂_t that passes through a point $x \in S$, then $\Sigma = \{I_x \cap S_{t_\Sigma(x)} | x \in S\}$. From this viewpoint, F can be regarded as a real-valued functional on $\{t_\Sigma\}$. Now if D_σ is compact, we can find the maximum and minimum values of t for the set D_σ to obtain an upper and lower bound on t_Σ that applies for all Σ. Moreover, the condition that Σ be compact and achronal ensures that $\{t_\Sigma\}$ is equicontinuous. These facts imply that $\mathcal{C}_\sigma$ is compact (with the uniform topology) and that the extreme value theorem applies to the function F.

In the case where D_σ is not compact (for instance, due to a singularity terminating a light sheet of σ), we can still argue that F has a maximum as long as D_σ satisfies a condition similar to but slightly stronger than the "compact restriction" idea discussed above. Suppose that B_+ is a surface in $\mathcal{C}_\sigma$ which is identical to Σ_+ in some neighborhood of S. For any $\Sigma \in \mathcal{C}_\sigma$, define another surface $\bar{\Sigma}$ by $t_{\bar{\Sigma}} = \min\{t_\Sigma, t_{B_+}\}$. If B_+ has the property that for any Σ we have $F(\Sigma) \leq F(\bar{\Sigma})$, then we will say that B_+ is a *future maximin barrier*. A past maximin barrier is defined analogously as a surface $B_- \in \mathcal{C}_\sigma$, identical to Σ_- in a neighborhood of S, such that for any Σ we have $F(\Sigma) \leq F(\bar{\Sigma})$ where $\bar{\Sigma}$ is defined by $t_{\bar{\Sigma}} = \max\{t_\Sigma, t_{B_-}\}$.

Now if D_σ possesses both a past and future maximin barrier, then we can restrict our attention to the subset of surfaces in $\mathcal{C}_\sigma$ that satisfy $t_{B_-} \leq t_\Sigma \leq t_{B_+}$. Let $\mathcal{C}_\sigma(B_-, B_+)$ denote this restricted set. Because B_- and B_+ are compact, $J_+(B_-) \cap J_-(B_+)$ is compact and so the set $\mathcal{C}_\sigma(B_-, B_+)$ is compact in the uniform topology and F has a maximum $\Sigma_0 \in \mathcal{C}_\sigma(B_-, B_+)$. The definition of past and future maximin barriers ensures us that if $\Sigma \in \mathcal{C}_\sigma$, then $F(\Sigma_0) \geq F(\Sigma)$. Thus, Σ_0 is a global maximum for F and we can safely define $\min(\partial A, \Sigma_0)$ as the maximin surface of A, $\text{Mm}(A)$.

As in the case of the compact restriction of D_σ used in theorem 1, it is difficult to find examples where D_σ does not possess a past and future barrier. Wall [187] argued that such barriers protect maximin surfaces from a wide range of singularities: approximately Kasner singularities, BKL singularities, and FRW big bangs all lead to past or future maximin barriers. If $\Sigma_\pm$ simply terminate at caustics rather than singularities, then $B_\pm = \Sigma_\pm$ are barriers. In any event, if $B_\pm$ exist, then the region $J_+(B_-) \cap J_-(B_+)$ provides a compact restriction of D_σ in the sense of theorem 1. Thus, the existence of $B_\pm$ ensures both the existence of $\text{Mm}(A)$ as well as the existence of ext (A). From here on, we will simply take for granted that a past and future maximin barrier exist.

Equivalence of Mm(A) and ext (A)

Below we will argue that $\mathrm{Mm}(A) = \mathrm{ext}\,(A)$. However, it is first very useful to introduce two additional definitions first.

1. Take $\Sigma \in \mathcal{C}_\sigma$ and let Γ be a codimension 2 surface anchored to ∂A that lies in D_σ. Consider the intersection between Σ and the future and past-directed orthogonal null surfaces of Γ that are directed toward A. This intersection is called the representative of Γ on Σ and will be denoted by $\mathrm{rep}(\Gamma, \Sigma)$.

2. The domain of dependence of codimension 1 achronal surfaces anchored to $A \cup \mathrm{ext}\, A$ lying in D_σ will be called the entanglement wedge of A.

Note that $\mathrm{rep}(\Gamma, \Sigma)$ is itself a codimension 2 surface anchored to ∂A that lies on Σ. Moreover, if Γ is extremal then, by the focusing theorem, $\mathrm{area}(\mathrm{rep}(\Gamma, \Sigma)) \leq \mathrm{area}(\Gamma)$.

We now demonstrate that our maximin procedure always finds $\mathrm{ext}\, A$, the extremal surface of minimal area that is anchored to ∂A and which lies in D_σ. While much of the proof is similar to the arguments in [187], we will have to pay special attention to the possibility that the maximin surface could run into the boundary of D_σ.

Theorem 2. *$Mm(A) = ext\,(A)$.*

Proof. The argument of theorem 15 in [187] immediately shows that if a point $p \in \mathrm{Mm}(A)$ is also in the interior of D_σ (i.e. $D_\sigma \setminus \partial D_\sigma$), then $\mathrm{Mm}(A)$ is extremal at p. In particular, if $\mathrm{Mm}(A) \cap \partial D_\sigma = \partial A$, then $\mathrm{Mm}(A)$ is an extremal surface everywhere. We now argue that $\mathrm{Mm}(A)$ in fact cannot ever intersect ∂D_σ outside of ∂A.

Suppose there exists $p \in \mathrm{Mm}(A) \cap (\partial D_\sigma \setminus \partial A)$. There must be an open neighborhood of p in $\mathrm{Mm}(A)$ (open in the d-2 dimensional manifold $\mathrm{Mm}(A)$) that is entirely contained in ∂D_σ. If this were not the case, $\mathrm{Mm}(A)$ would be extremal at points arbitrarily close to p and would thus be extremal at p. Moreover, $\mathrm{Mm}(A)$ would be tangent to ∂D_σ at p. However, ∂D_σ is an extremal surface barrier (see lemma 1) so this is not possible. There are now two cases to consider.

- *Case 1*: $p \in \partial D_\sigma \setminus \sigma$.
  Figure 5.5 illustrates a construction that we will use for this case. Take $p \in \Sigma_+$ (the case of $p \in \Sigma_-$ is no different). By construction, $\mathrm{Mm}(A)$ is minimal on a surface Σ_0. There exists a (dimension d-1) open subset U of Σ_0 containing p such that $U \cap \mathrm{Mm}(A) \subset \Sigma_+$. Moreover, we can require that U is "split" by $\mathrm{Mm}(A)$ into two disconnected sets, N and V, such that N is the side of U closer to σ. Since Σ_0 is anchored to σ, we must have that $N \subset \Sigma_+$ and, in particular, N is null. On the other hand, V cannot be a subset of Σ_+. If it were, then $\mathrm{Mm}(A)$ could decrease its area by being deformed up Σ_+ (by the focusing theorem). In particular, we can take U small enough to ensure that V is nowhere null in the direction of k.

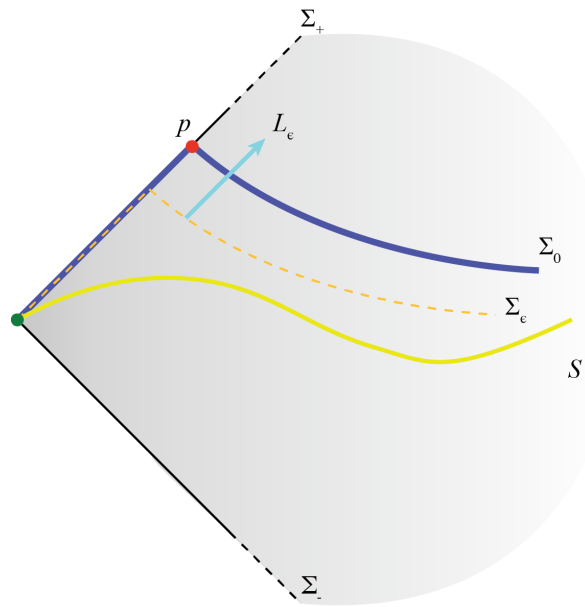


Figure 5.5: This figure depicts the argument of case 1 of the proof of theorem 2. Note that the surface S is shown here for reference and that it does not play a critical role in the proof. The shaded region is $D_\sigma = D(S)$ and the green dot is (a cross-section of) the leaf σ.

We now consider the process of slightly sliding Σ_0 down Σ_+. More precisely, take a small parameter $\epsilon > 0$ and a corresponding one-parameter family of slices $\{\Sigma_\epsilon\}$ that are slightly deformed from Σ_0 in a way we now describe (an example of Σ_ϵ is depicted in figure 5.5 by an orange dashed line). The surface $\mathrm{Mm}(A) \cap U$ is described by a function $\lambda_0(x)$ giving the affine distance from σ up to $\mathrm{Mm}(A)$ at a point $x \in \sigma$. Now put $\lambda_\epsilon(x) = \lambda_0(x) - \epsilon f(x)$. Here, $f : \sigma \to [0, 1]$ is a smooth weighting function which equals 1 at the null generator x_p that p lies on. We take f to go to zero smoothly as x moves away from x_p, equaling zero exactly when x corresponds to a point outside of $U \cap \mathrm{Mm}(A)$. For $\lambda < \lambda_\epsilon(x)$, we require that the surface Σ_ϵ is identical to Σ_+. We extend Σ_ϵ beyond λ_ϵ by parallel transporting tangent vectors on $\mathrm{Mm}(A)$ directed toward V down to λ_ϵ. This prescription does not uniquely fix Σ_ϵ, but it is sufficient for our purposes.

Consider the one-parameter family of codimension 2 curves $\min(\partial A, \Sigma_\epsilon)$. For any $\epsilon > 0$, let L_ϵ denote the future-directed null congruence of $\min(\partial A, \Sigma_\epsilon)$ that points toward the interior of D_σ (see figure 5.5). The continuity of $\min(\partial A, \Sigma_\epsilon)$ as ϵ varies and the fact that Σ_+ is a light sheet ensures that there exists and $\epsilon_0 > 0$ such that for $\epsilon < \epsilon_0$,

- L_ϵ intersects Σ_0 to form a codimension 2 surface on Σ_0 anchored to ∂A and
- L_ϵ has negative future-directed expansion in the region between $\min(\partial A, \Sigma_\epsilon)$ and its intersection with Σ_0.

Denote this intersection by C_ϵ and observe that $C_0 = \mathrm{Mm}(A)$. But $\mathrm{Mm}(A)$ is minimal on Σ_0 so for sufficiently small ϵ,

$$\mathrm{area}(\mathrm{Mm}(A)) < \mathrm{area}(C_\epsilon) \leq \mathrm{area}(\min(\partial A, \Sigma_\epsilon))$$

which contradicts the assumption that $\mathrm{Mm}(A)$ has area greater than or equal to the minimal area surface on any slice. Note that the last inequality above follows from the focusing theorem applied to L_ϵ.

- *case 2*: $p \in \sigma$.
  Assume that there exists a (dimension d-2) open subset of $\mathrm{Mm}(A)$ that is contained in σ. (If not, there must be such an open set in $\partial D_\sigma \setminus \sigma$ which just leads to case 1 above.) Now consider the null vector field k on σ and the geodesics generated by it. Follow these geodesics from σ up along Σ_+ by a short affine distance $\epsilon > 0$ to generate a new codimension 2 surface, σ_ϵ, which limits to σ when $\epsilon \to 0$. The focusing theorem now gives rise to a modified version of equation 5.1 at σ_ϵ:

$$\begin{aligned} \theta^l_\epsilon &> 0 \\ \theta^k_\epsilon &< 0. \end{aligned} \tag{5.5}$$

  Along with moving σ up the light-sheet, we also translate A up the sheet to a one-parameter family of surfaces A_ϵ that limit to A. Consider the maximin construction applied to codimension 2 surfaces anchored to ∂A_ϵ that lie on codimension 1 surfaces anchored to σ_ϵ. We denote the result by $\mathrm{Mm}(A_\epsilon)$. We also define D_{σ_ϵ} in the obvious way. Now this maximin procedure leads to the same two cases that we are now studying. The first case, where $\mathrm{Mm}(A_\epsilon)$ intersects $\partial D_{\sigma_\epsilon} \setminus \sigma_\epsilon$ proceeds exactly as it did with $\epsilon = 0$. Now suppose that $\mathrm{Mm}(A_\epsilon)$ has an open set contained in σ_ϵ. $\mathrm{Mm}(A_\epsilon)$ must be minimal on some slice Σ_ϵ. However, equation 5.5 implies that σ_ϵ has negative (inward) extrinsic curvature on Σ_ϵ. It is thus impossible for $\mathrm{Mm}(A_\epsilon)$ to be minimal on Σ_ϵ since its area could be decreased by "cutting corners."

  We can thus conclude that $\mathrm{Mm}(A_\epsilon) \cap \partial D_{\sigma_\epsilon} = \partial A_\epsilon$. This implies that $\mathrm{Mm}(A_\epsilon)$ is extremal. Taking the limit as $\epsilon \to 0$, we conclude that $\mathrm{Mm}(A)$ is extremal. But, given our assumption that part of $\mathrm{Mm}(A)$ lies on σ, equation 5.1 shows that $\mathrm{Mm}(A)$ cannot be extremal since extremal surfaces have zero null expansion in all directions.

At this point it is proven that $\mathrm{Mm}(A)$ is extremal. All that is left is to show that, of all the extremal surfaces in D_σ that are anchored to ∂A, $\mathrm{Mm}(A)$ is the smallest. Let $\Sigma_0 \in \mathcal{C}_\sigma$ be a slice on which $\mathrm{Mm}(A)$ is minimal. If Γ is another extremal surface anchored to ∂A then, as a result of the focusing theorem, we find that

$$\mathrm{area}(\mathrm{Mm}(A)) \leq \mathrm{area}(\mathrm{rep}(\Gamma, \Sigma_0)) \leq \mathrm{area}(\Gamma).$$

□

We are now in a position to prove a variety of properties of screen entanglement entropy. We begin with the "Page bound" advertised in section 5.2.

Corollary 1. *If A is a region in the leaf σ, then*

$$S(A) \leq \min\{S_{extensive}(A), S_{extensive}(\sigma \setminus A)\}$$

where S deonotes the holographic screen entanglement entropy of A and $S_{extensive}(X)$ denotes the area of a region $X \subset \sigma$ divided by 4.

Proof. $S(A) = \text{area}(\text{Mm(A)})/4$ but $\text{Mm}(A) = \min(\partial A, \Sigma_0)$ for some $\Sigma_0 \in \mathcal{C}_\sigma$. Both A and $\sigma \setminus A$ are codimension $d-2$ dimensional surfaces on Σ_0 anchored to ∂A so the area of $\text{Mm}(A)$ is less than or equal to the areas of both A and $\sigma \setminus A$. □

Next we turn to the proof of strong subadditivity for holographic screen entanglement entropy (other properties of entanglement entropy that admit covariant geometrical bulk proofs can be imported here as well). Unlike the case of theorems 1 and 2, the arguments below are essentially identical to those of [187] with little additional subtlety. We start with our version of theorem 17 in [187] which states that if $B \subset A$, then ext A lies "outside" of ext B.

Theorem 3. *Suppose that A and B are regions in the leaf σ with $B \subset A$. Then,*

1. *the entanglement wedge of A contains the entanglement wedge of B,*
2. *there exists a surface in $\mathcal{C}_\sigma$ on which both ext A and ext B are minimal.*

Sketch of Proof: The proof is the same as that of theorem 17 of [187] so we only sketch it here. For any surface in $\Sigma \in \mathcal{C}_\sigma$, consider a pair of codimension 2 surfaces constrained to lie on Σ, Γ_A and Γ_B, such that Γ_A is anchored to ∂A and Γ_B is anchored to ∂B. Then let $Z = \text{area}(\Gamma_A) + \text{area}(\Gamma_B)$. We now minimize the value of Z by varying over all possible choices of Γ_A and Γ_B. After that, we maximize the minimal values of Z by varying over all possible Σ.

This new maximin procedure gives a well-defined answer for the maximinimal value of Z. Moreover, a slice Σ_0 results on which both Γ_A and Γ_B are minimal. On this slice, it is impossible for Γ_A to cross Γ_B as this would necessarily give rise to a surface on Σ_0 anchored to ∂A with smaller area than Γ_A. A further observation is that if a connected component of A is distinct from a component of B, the corresponding connected components of Γ_A and Γ_B cannot come into contact even tangentially. The argument for this is that the component of Γ_B would necessarily have a different trace of its spatial extrinsic curvature than Γ_A at points close to the contact point. This would mean that either Γ_A or Γ_B is not minimal on Σ_0.

At this point it is known that components of Γ_A or Γ_B that are distinct have neighborhoods in Σ_0 that do not intersect the other surface. Within such neighborhoods, small

deviations Σ_0 and the minimal surfaces can be made that prove that such surfaces are extremal.

The only remaining step is to show that, in fact, Γ_A and Γ_B are the extremal surfaces in D_σ of minimal area. If Γ'_A is an extremal surface in D_σ anchored to ∂A, then its representation on Σ_0 must have larger area than that of Γ_A but smaller area than that of Γ'_A. Thus, $\Gamma_A = \text{ext } A$. Similarly, $\Gamma_B = \text{ext } B$. By construction, both are minimal on the same surface $\Sigma_0 \in \mathcal{C}_\sigma$. Moreover, because Σ_0 is achronal, we must have that the entanglement wedge of A contains that of B. □

Corollary 2. *Suppose that A, B, and C are nonintersecting regions in σ. Then,*

$$S(AB) + S(BC) \geq S(ABC) + S(B)$$

where XY denotes $X \cup Y$ and where the function S is defined in equation 5.2.

Proof. By theorem 3, we can find a surface $\Sigma_0 \in \mathcal{C}_\sigma$ such that ext B and ext ABC are both minimal on Σ_0. Let $\tilde{S}(AB)$ and $\tilde{S}(BC)$ denote the areas of the representations of ext AB and ext BC on Σ_0. Then,

$$S(AB) + S(BC) \geq \tilde{S}(AB) + \tilde{S}(BC) \geq S(ABC) + S(B)$$

where the first inequality follows from the focusing theorem and the second inequality follows from the standard geometric proof of strong subadditivity [92]. □

Note that the inequality $S(A) + S(B) \geq S(AB)$ follows as a special case of this result.

5.4 Extremal Surfaces in FRW Cosmology

The conventional holographic entanglement entropy prescription, with its limitation to asymptotically locally AdS spacetimes, provides very little information about entanglement structure in cosmology. One of the most intriguing applications of our proposal, therefore, is to calculate holographic screen entanglement entropy in FRW universes. Assuming the screen entanglement conjecture, the calculations below give the entanglement entropy of subsystems in quantum states that are dual to cosmological spacetimes.

Holographic Screens in FRW Cosmology

First we review the holographic screen structure of FRW spacetimes. Consider a homogeneous and isotropic spacetime with the metric

$$ds^2 = -d\tau^2 + a(\tau)^2 \left(d\chi^2 + f(\chi)^2 d\Omega_2^2\right) \tag{5.6}$$

where $f(\chi) = \sinh(\chi), \chi$, or $\sin(\chi)$ in the open, flat, and closed cases respectively. Before computing extremal surfaces we must decide upon a null foliation for the spacetime and then

identify the corresponding holographic screen. Null foliations (and thus holographic screens) are highly nonunique. The foliation we will consider here is that of past light cones from a worldline at $\chi = 0$.

To find the holographic screen for this foliation, it is convenient to introduce a conformal time coordinate η such that $d\tau/d\eta = a$. Then, the past light cone of the point $(\eta = \eta_0, \chi = 0)$ satisfies $\chi = \eta_0 - \eta$. Spheres along the past light cone can be parameterized by the coordinate η, and their area is given by

$$\mathcal{A}(\eta) = 4\pi\, a\Big(\tau(\eta_0 - \eta)\Big)^2 f(\eta_0 - \eta)^2 \tag{5.7}$$

Assuming that $a = 0$ is not merely a coordinate singularity, the condition that $\mathcal{A}$ is maximized is equivalent to the condition that $d\mathcal{A}/d\eta = 0$. Thus, equation 5.7 gives the condition that fixes the holographic screen:

$$\frac{f(\chi)}{f'(\chi)} - \frac{1}{\dot{a}(\tau)} = 0. \tag{5.8}$$

The codimension 1 surface defined by this constraint may be timelike, spacelike, or null, depending on the particular choice of FRW spacetime. The foliating leaves of this holographic screen are spheres of constant τ and comoving radius χ satisfying equation 5.8. The covariant entropy bound implies that each leaf has sufficient area to holographically encode the information on one past light cone from the worldline at $\chi = 0$ [27, 28].

Let $\sigma(\tau)$ be the leaf of the holographic screen at time τ and let $\rho(\tau)$ denote the energy density in the universe (measured by comoving observers) at time τ. Then, one can write a simple expression for the area of a leaf of the holographic screen at time τ, valid for any f:

$$\text{area}(\sigma(\tau)) = \frac{3}{2\rho(\tau)}. \tag{5.9}$$

In particular, this expression shows that holographic screens grow in area as the universe expands.

Extremal Surfaces in de Sitter Space

Consider 3+1 dimensional de Sitter space of radius α. This spacetime is $S^3 \times \mathbf{R}$ with the metric

$$ds^2 = -dT^2 + \alpha^2 \cosh^2\left(\frac{T}{\alpha}\right) d\Omega_3^2$$

where $d\Omega_3^2$ is the metric on a unit 3-sphere. Despite the fact that this spacetime has the form of equation 5.6 (with $f(\chi) = \sin\chi$), it is an awkward setting for the consideration of holographic screens: the null expansion on the past or future light cones of any point in de Sitter space goes to zero only at infinite affine parameter. This suggests that the appropriate "boundary" of de Sitter space is past or future infinity. Even if we do attempt to anchor

extremal surfaces to spheres at infinity, the analysis in section 5.3 fails to apply because of the assumption made there that leaves are compact.

Fortunately these difficulties can be averted completely by considering an FRW spacetime that asymptotically approaches de Sitter space at late times. Specifically, we will consider a spacetime of the form of equation 5.6 with vacuum energy density ρ_Λ and, in addition, some matter content $\rho_{\text{matter}}(\tau)$ with the property the matter content gives rise to a big bang at $\tau = 0$ and dilutes completely[12] as $\tau \to \infty$.

Equation 5.9 immediately implies that

$$\lim_{\tau\to\infty} \text{area}(\sigma(\tau)) = \frac{3}{2\rho_\Lambda} = 4\pi\alpha^2 \tag{5.10}$$

where $\alpha = \sqrt{3/8\pi\rho_\Lambda}$. Because of the big bang singularity, we must have that $\text{area}(\sigma_{\tau=0}) = 0$. Thus, by the area law for holographic screens [31, 32], we can conclude that the leaves of our screen are spheres that monotonically increase in area, starting with 0 area at the big bang, and expanding to approach the de Sitter horizon of area $4\pi\alpha^2$ at late τ.

Now focus on a late time leaf $\sigma(\tau)$. As discussed in section 5.3, given a region $A \subset \sigma(\tau)$ with a boundary, we can determine the holographic screen entanglement entropy of A, $S(A)$, by considering an extremal surface anchored to and terminating at ∂A. In the notation of section 5.3, $D_{\sigma(\tau)}$ is compact so theorem 1 implies that an extremal surface anchored to ∂A exists and lies inside of $D_{\sigma(\tau)}$.

For any time τ, define

$$S^\tau_{\text{Page}}(A) = \begin{cases} \frac{1}{4}\text{area}(A) & \text{area}(A) \leq \frac{1}{2}\text{area}(\sigma(\tau)) \\ \frac{1}{4}\left(\text{area}(\sigma(\tau)) - \text{area}(A)\right) & \text{area}(A) > \frac{1}{2}\text{area}(\sigma(\tau)). \end{cases} \tag{5.11}$$

We allow this definition to extend to a function $S^\infty_{\text{Page}}(A)$ where A is a region on a 2-sphere of radius α. This $\tau = \infty$ case is defined exactly as in equation 5.11 if we take $\text{area}(\sigma(\infty)) = 4\pi\alpha^2$.

Below we will present an argument that if $A \subset \sigma(\tau)$, then

$$\lim_{\tau\to\infty} S(A) = S^\infty_{\text{Page}}(A). \tag{5.12}$$

(Note that in this limit, it is implied that A is transported to later and later leaves.) Thus, we will find that as $\tau \to \infty$, $S(A)$ approaches the random entanglement limit discussed in section 5.2.

Any interpretation of this result is necessarily speculative. Nevertheless, if one assumes the screen entanglement conjecture, then equation 5.12 implies that the the quantum state of an FRW universe asymptotically approaching de Sitter space has the property that its $O(\alpha^2)$ degrees of freedom are almost randomly entangled with one-another. At earlier times, the degrees of freedom are not randomly entangled because $S(A) < S^\infty_{\text{Page}}(A)$.

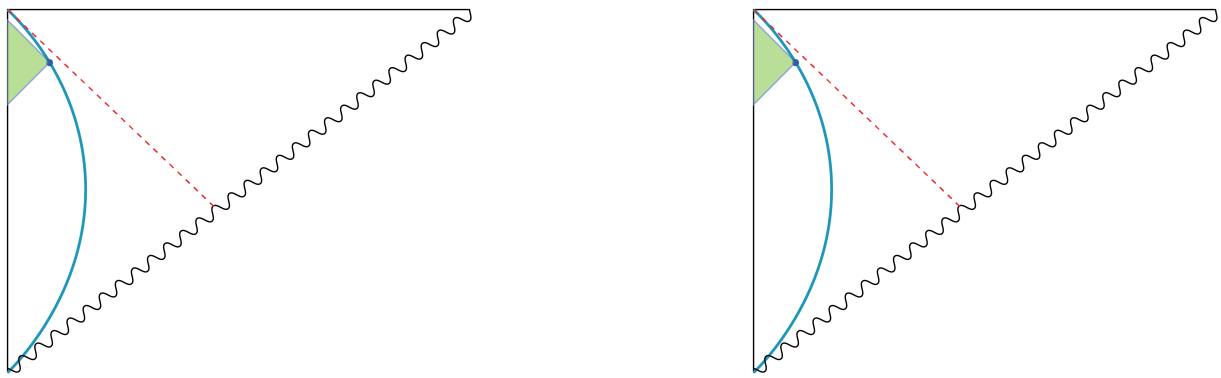


Figure 5.6: The domain of dependence $D_{\sigma(\tau)}$ for a late time leaf in the flat FRW universe (the small green triangle in the upper diagram) can be approximately mapped to a domain of dependence $D_{\tilde{\sigma}(\tau)}$ in empty de Sitter space (lower diagram). The mapping becomes increasingly accurate as τ becomes larger. The effect of increasing τ is to move the green triangle in the upper diagram into the top-left corner (along the blue curve), while the green triangle in the lower diagram moves to the right and approaches the entire left static wedge.

Random Entanglement and the Static Sphere Approximation

We now present a combination of rigorous arguments, numerical data, and analytic approximations suggesting that the approximate de Sitter cosmological spacetimes discussed above saturate the random entanglement bound in the $\tau \to \infty$ limit. As before, $\sigma(\tau)$ denotes a leaf at time τ in an FRW universe with vacuum energy as well as matter energy that dilutes at late time.

The entire region $D_{\sigma(\tau)}$ has a metric that can be made arbitrarily similar to that of a patch of empty de Sitter space by making τ large. To see this, first note that points in $D_{\sigma(\tau)}$ have $\chi < \chi_{\text{screen}}(\tau)$ and $\chi_{\text{screen}}(\tau)$ can be made arbitrarily small by making τ large. (This follows from equation 5.10 and the fact that $\lim_{\tau\to\infty} a(\tau) = \infty$.) Meanwhile, the conformal diagram for our spacetime immediately shows that the minimal value of τ in $D_{\sigma(\tau)}$ can be made arbitrarily large by making τ large. Thus $D_{\sigma(\tau)}$ can be made to only cover arbitrarily large τ and arbitrarily small χ, in which case our metric of equation 5.6 takes the form

$$ds^2 \approx -d\tau^2 + c\, e^{2\tau/\alpha}(d\chi^2 + \chi^2 d\Omega_2^2) \tag{5.13}$$

where c is a constant and α is the same constant as before. Here we have made use of the Friedmann equations. The right-hand side of this equation is precisely the metric of de Sitter space in flat slicing. De Sitter space can also be described in static coordinates that make a time-translation Killing vector field manifest:

$$ds^2 \approx -\left(1 - \frac{r^2}{\alpha^2}\right) dt^2 + \left(1 - \frac{r^2}{\alpha^2}\right)^{-1} dr^2 + r^2 d\Omega_2^2. \tag{5.14}$$

[12] In particular, we are not considering spacetimes with a big crunch in this section.

Fortunately, $D_{\sigma(\tau)}$ lies in a region that is well-described by either the flat or static slicing of equations 5.13 and 5.14 respectively.

We can now identify $D_{\sigma(\tau)}$ with a region $D_{\tilde{\sigma}(\tau)}$ where $D_{\tilde{\sigma}(\tau)}$ denotes a corresponding region in exact de Sitter space obtained by finding a sphere $\tilde{\sigma}(\tau)$ in the static patch with area matching that of $\sigma(\tau)$. While it may seem natural to put $\tilde{\sigma}(\tau)$ at large static time, we can use the t translational symmetry of de Sitter space to place $\tilde{\sigma}(\tau)$ at $t = 0$ for all τ. The effect of increasing τ is simply to bring $\tilde{\sigma}(\tau)$ closer to the bifurcation sphere on the de Sitter horizon. This identification is illustrated in figure 5.6. Note that as $\tau \to \infty$, the geometry of $D_{\sigma(\tau)}$ and $D_{\tilde{\sigma}(\tau)}$ become arbitrarily similar.

Consider the region $A \subset \sigma(\tau)$ which can be identified with a region $\tilde{A} \subset \tilde{\sigma}(\tau)$. At large τ, $\tilde{\sigma}(\tau)$ approaches the equator of a 3-sphere of radius α. The equator itself is an extremal surface so with $\tau < \infty$ but still large, there must be an extremal surface that is close to $\tilde{A}$ but not exactly on it. Its area will be slightly less than that of $\tilde{A}$. Note, moreover, that if the area of $\tilde{A}$ exceeds half the area of the equator, then a smaller extremal surface can be obtained by considering the complement of A.

This suggests but does not yet prove that at large τ, the holographic screen entanglement entropy of A is almost equal to a fourth of its own area in Planck units if A has less area than half of the de Sitter horizon. What we have proven so far is that an extremal surface exists with area almost equal to that of A (or $4\pi\alpha^2 - \text{area}(A)$).

What if there is another extremal surface with smaller area than the one we have found? It is easy to see that this is impossible. Following the notation in section 5.3, consider the spacelike surface Σ_0 that, after mapping to $D_{\tilde{\sigma}(\tau)}$, lies at static time $t = 0$, and that and terminates at $\tilde{\sigma}$. (Σ_0 is most of a hemisphere of the 3-sphere.) The Riemannian geometry of S^3 shows that the surface of minimal area anchored to ∂A is the one we have already found. If Γ is another extremal surface (not necessarily lying on Σ_0), then its representation on Σ_0, $\text{rep}(\Gamma, \Sigma_0)$, necessarily has larger area than the extremal surface close to the horizon. But $\text{area}(\Gamma) \geq \text{area}(\text{rep}(\Gamma, \Sigma_0))$ so we conclude that Γ does not have minimal area.

The arguments above show that the random entanglement limit is saturated at large τ. Taking $0 \ll \tau < \infty$ and $A \subset \sigma(\tau)$, we now explain a way to obtain a more accurate estimate for $S(A)$ than $S^{\tau}_{\text{Page}}(A)$. Calculating $S(A)$ without taking the large τ limit is more involved than what was done above. Nonetheless, it is worthwhile to investigate this case to better understand how the Page bound limit is approached. In particular, it is of interest to understand how the discontinuity of the derivative of S^{∞}_{Page} arises.

We begin by further discussing the role of the 3-sphere in de Sitter space. Figure 5.7 depicts a hemisphere of an S^3 of radius α which is precisely half of a static slice of de Sitter space (which we can freely take to be $t = 0$). Define a parameter z as $z = \sqrt{\alpha^2 - r^2}$ where r is the static radius appearing in equation 5.14. Note that a surface of constant z (and static time) is an S^2 of area $4\pi(\alpha^2 - z^2)$. This suggests a way to obtain an approximation for $S(A)$ if A is a region in the leaf $\sigma(\tau)$. Rather than taking A to be a region in $\sigma(\tau)$, we take figure

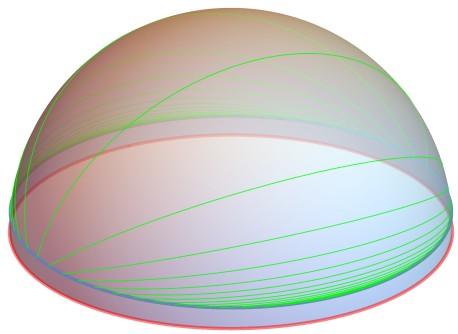


Figure 5.7: The upper hemisphere of a 3-sphere of radius α is half of a static slice in empty de Sitter space and serves as a good approximation for $D_{\sigma(\tau)}$ at large τ. The blue 2-sphere (appearing as a circle here) lies at constant z (equivalently, constant r where r is the radial coordinate in equation 5.14). This 2-sphere is an approximation for the leaf $\sigma(\tau)$. Green surfaces depict extremal spherical caps on S^3 that approximate ext A_ψ for various values of ψ. The many samples of extremal surfaces shown here have evenly spaced values of ψ. Figure 5.8 provides evidence that this static sphere approximation is accurate at late τ.

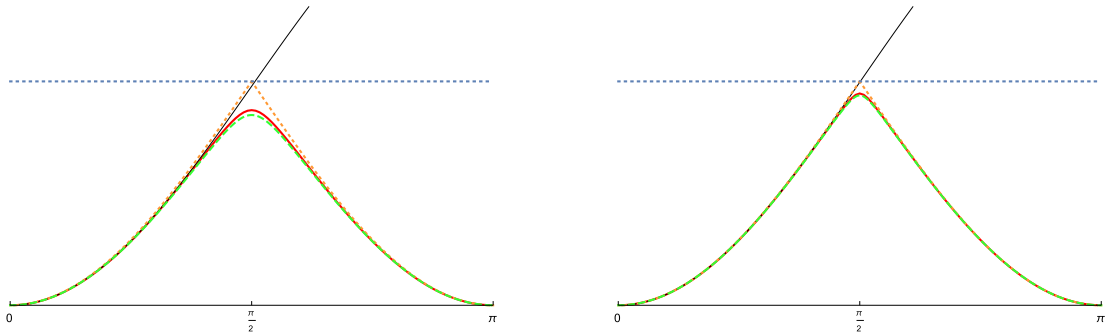
5.6 seriously and map A to a region in the S^2 of constant

$$z = \sqrt{\frac{4\pi\alpha^2 - \text{area}(\sigma(\tau))}{4\pi}} \tag{5.15}$$

which ensures that this S^2 has the same area as $\sigma(\tau)$. After this mapping is made, one computes $S(A)$ by finding the extremal surface on the S^3 that is anchored to ∂A (which we take to lie at constant z). Below we will refer to this procedure as the "static sphere approximation."

Consider regions in $\sigma(\tau)$ that are spherical caps. Such a cap can be fixed (up to $SO(3)$ rotation) by a zenith opening angle angle ψ, so we will denote our region of $\sigma(\tau)$ by A_ψ. (With this notation, $A_{\pi/2}$ is a hemisphere and A_π is the entire leaf.) The static sphere approximation makes it is clear that for $0 < \psi \ll \pi/2$, ext A_ψ is close to A_ψ itself and that for $\pi/2 \ll \psi < \pi$, ext A_ψ approaches $\sigma(\tau) \setminus A_\psi$. As ψ passes the transition angle $\pi/2$, ext A_ψ quickly passes over the top of the 3-sphere of radius α. The closer area$(\sigma(\tau))$ is to $4\pi\alpha^2$, the faster ext A_ψ passes over the top of the sphere. This explains how the discontinuity in the derivative of $S^\infty_{\text{Page}}(A)$ arises in the large τ limit.[13]

[13]For finite τ, there is always another extremal surface on the 3-sphere which goes around the sphere the wrong way. This surface always has area greater than ext A_ψ and, in any case, fails to lie in $D_{\sigma(\tau)}$. However, if we consider the $\tau = \infty$ limit, then ext A_ψ does not smoothly pass over the hemisphere of the 3-sphere,



(a) $z/\alpha \approx .05$ (b) $z/\alpha \approx .02$

Figure 5.8: Plots of $S(A_\psi)$ and other quantities for two leaves at different times in a universe with dust and vacuum energy. In both plots, the red curve is the numerically computed holographic screen entanglement entropy of A_ψ. The dashed green curve is the static sphere approximation for $S(A\psi)$ which becomes more accurate at later τ (smaller z). The orange curve with a sharp peak is $S_{\text{Page}}(A_\psi)$ as defined by equation 5.11 and the black curve is $S_{\text{extensive}}(A_\psi)$. The horizontal line, provided for scale, marks the value of $\pi\alpha^2/2$ which is precisely one fourth of the extensive entropy of the de Sitter horizon.

Because the geometry of S^3 is simple, it is not difficult to obtain an explicit (if cumbersome) expression for $S(A_\psi)$ in the static sphere approximation:

$$S(A_\psi) \approx \pi \sin^2\left(\frac{1}{4}\cos^{-1}\left[\frac{z^2}{\alpha^2} + \left(1 - \frac{z^2}{\alpha^2}\right)\cos 2\psi\right]\right) \tag{5.16}$$

where z is given by equation 5.15 and, as before, $\alpha = \sqrt{3/8\pi\rho_\Lambda}$. This expression can be thought of as giving a correction to the "zeroth order" expression $S(A_\psi) \approx S^\infty_{\text{Page}}(A_\psi)$. Taking $\tau < \infty$ will lead to corrections in $1/\tau$ that are not described by the static sphere method. It is an open question as to whether or not such corrections can, in principle, be of the same (or greater) order in $1/\tau$ as the one we have studied here. However, numerical data that suggests that the static sphere approximation is accurate at large τ as we will now see.

As explained above, the cosmological spacetimes we have been discussing have vacuum energy ρ_Λ as well as some matter content that dilutes at late time. The simplest case of this is when the universe is flat ($f(\chi) = \chi$) and when the additional matter content consists of only one species with density ρ_{matter} and pressure $p_{\text{m}} = w\rho_{\text{m}}$. The scale factor for this case is

$$a(\tau) = C \sinh\left[\frac{3(1+w)\tau}{2\alpha}\right]^{\frac{2}{3(1+w)}} \tag{5.17}$$

and in this case, the discontinuity in the derivative of $S^\infty_{\text{Page}}(A)$ is explained by the fact that the surface that wraps around the sphere the "wrong way" is now precisely the complement of A_ψ in the equator. If ψ exceeds $\pi/2$ in this case, then the complement of A_ψ has smaller area than A_ψ. We see that a phase transition occurs only in the exact $\tau = \infty$ limit.

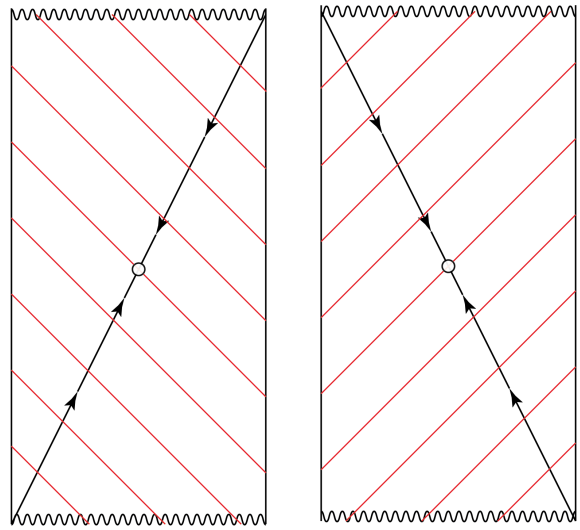


Figure 5.9: Both Penrose diagrams here are for the same spacetime: a closed FRW universe with dust. The red lines denote a null foliation and the black diagonals are the past and future holographic screens corresponding to the foliation. The two figures demonstrate that different foliations give rise to different screens. In both figures, the lower half of the diagonal is a past holographic screen and the upper half is a future holographic screen. Arrows show the direction of increasing area.

where the normalization factor C is independent of τ.

This setting is very useful to test the theoretical apparatus developed in this section. In the case of $w = 0$, figure 5.8 shows a variety of quantities we have discussed. Figure 5.8 (a) and (b) depict the case of an earlier and later time leaf with $z/\alpha \approx .05$ and $z/\alpha \approx .02$ respectively. The solid red curves show $S(A_\psi)$ (computed numerically) while the green curves give the static sphere approximation of equation 5.16. The dotted horizontal line marks half of the de Sitter entropy: $S_{1/2} = \pi\alpha^2/2$. As expected, $S(A_\psi) < S_{1/2}$. The orange curve with a discontinuity in its derivative is $S^{\infty}_{\text{Page}}(A_\psi)$. Comparing figures 5.8 (a) and (b), one can see that $S(A_\psi)$ is approaching $S^{\infty}_{\text{Page}}(A_\psi)$ as $\tau \to \infty$. Finally, the black curves shows extensive entropy: $S_{\text{extensive}}(A_\psi) = (1/4)\text{area}(A_\psi)$. Note that $S(A_\psi) < S_{\text{extensive}}(A_\psi)$ for all ψ as required by corollary 1.

Closed Universe with a Big Crunch

The holographic screen entanglement entropy structure of a closed universe with a past and future singularity is similar to that of approximate de Sitter space. The spacetimes we consider have the metric of equation 5.6 with $f(\chi) = \sin(\chi)$. In this case the coordinate χ takes values from 0 to π. We put one species of matter content in the spacetime that

satisfies $p = w\rho$ which gives rise to a big bang at $\tau = 0$ as well as a big crunch. As before, we introduce a conformal time coordinate η in terms of which the scale factor is

$$a(\eta) = c\left(\sin\frac{\eta}{q}\right)^q$$

where $q = 2/(1+3w)$ and c is constant. This shows that the Penrose diagram for this spacetime is a rectangle with a time-to-space aspect ratio of q.

Figure 5.9 shows the holographic screen structure of this spacetime for two examples of null foliations. We focus on the diagram to the left in which case the null foliation (partially) consists of past light cones of a comoving worldline at the $\chi = 0$. As suggested by the figure, the holographic screen is given by

$$\chi_{\text{screen}} = \frac{1}{q}\eta.$$

However, a subtlety arises because the screen is a past holographic screen for $\eta < q\pi/2$ and a future screen for $\eta > q\pi/2$. The sphere that connects the past and future screen is extremal (this was called an "optimal" surface in [28]) and has area $4\pi c^2$. Let $\sigma(\eta)$ denote the leaf at conformal time η. We put $\sigma_0 = \sigma(\eta = q\pi/2)$.

Just as in the de Sitter case, this example leads to a saturation of the Page bound of equation 5.3 as leaves are maximized in area. More precisely, if $A \subset \sigma(\eta)$, then $\lim_{\eta\to q\pi/2} S(A) = S^{\infty}_{\text{Page}}(A)$ where in this case

$$S^{\infty}_{\text{Page}}(A) = \begin{cases} \frac{1}{4}\text{area}(A) & \text{area}(A) \leq \frac{1}{2}\pi c^2 \\ \frac{1}{4}\left(4\pi c^2 - \text{area}(A)\right) & \text{area}(A) > \frac{1}{2}\pi c^2. \end{cases}$$

It appears that $S(A)$ saturates the Page bound in a great variety of cases where the areas of leaves are bounded above.

5.5 Concluding Remarks

The proposal we have given above may open the door to a new research program: the study of the entanglement structure of general spacetimes. In light of this, and for the sake of clarity, we now summarize the recipe for computing von Neumann entropy under the assumption of the screen entanglement conjecture discussed in section 5.2:

1. Select a particular null foliation $\{N_r\}$ of a spacetime with dimension d.
2. Find the codimension 2 surfaces $\{\sigma_r\}$ with $\sigma_r \subset N_r$ that have maximal area on each N_r.
3. Take a $d-2$ dimensional subregion $A \subset \sigma_r$ with a boundary ∂A.

4. Of all extremal surfaces anchored to ∂A and lying in the causal region D_σ (see section 5.2), select the one of minimal area. The conjectured entropy $S(A)$ is then one fourth the area of the minimal extremal surface in Planck units.

Potential applications of our conjecture are numerous. One example not considered above is case of a spacetime with a black hole. Black holes formed through the collapse of matter possess future holographic screens in their interiors that approach their horizons at late times. It is of potential significance to investigate the entanglement structure of such spacetimes. Perhaps such an analysis will shed light on the firewall paradox [10].

If the screen entanglement conjecture is correct, it should still only be regarded as a leading order prescription for the computation of von Neumann entropies. A version of the analysis of [64] may be extendible to the context of holographic screens. It is not completely obvious how this should be done. If A is a region in a leaf σ lying on a Cauchy slice S_0, one may consider the region on S_0 bounded by A and its extremal surface ext (A) and compute the entanglement entropy of this region in a quantum field theory on the spacetime background. On the other hand, it may be necessary to modify the spacetime position of the holographic screen itself as was done in [30].

Chapter 6

Toward a Holographic Theory for General Spacetimes

6.1 Introduction

As with any other classical object, spacetime is expected to consist of a large number of quantum degrees of freedom. The first explicit hint of this came from the discovery that empty spacetime can carry entropy [21, 22, 20, 84, 86, 70]. What theory describes these degrees of freedom as well as the excitations on them, i.e. matter?

Part of the difficulty in finding such a theory is the large redundancies present in the description of gravitational spacetime. The holographic principle [173, 169, 29] suggests that the natural space in which the microscopic degrees of freedom for spacetime (and matter) live is a non-dynamical spacetime whose dimension is one less than that in the original description (as demonstrated in the special case of the AdS/CFT correspondence [122]). This represents a huge redundancy in the original gravitational description beyond that associated with general coordinate transformations. For general spacetimes, causality plays a central role in fixing this redundancy [65, 27]. A similar idea also plays an important role in addressing problems in the semiclassical descriptions of black holes [170] and cosmology [132, 35].

In this paper, we explore a holographic theory for general spacetimes. We follow a "bottom-up" approach given the lack of a useful description in known frameworks, such as AdS/CFT and string theory in asymptotically Minkowski space. We assume that our holographic theory is formulated on a holographic screen [28], a codimension-1 surface on which the information about the original spacetime can be encoded. This construction can be extended beyond the semiclassical regime by considering all possible states on all possible slices—called leaves—of holographic screens [132, 133], where the nonuniqueness of erecting a holographic screen is interpreted as the freedom in fixing the redundancy associated with holography. The resulting picture is consistent with the recently discovered area theorem applicable to the holographic screens [31, 32, 159].

To study the structure of the theory, we use conjectured relationships between spacetime in the gravitational description and quantum entanglement in the holographic theory. Recently, it has become increasingly clear that quantum entanglement among holographic degrees of freedom plays an important role in the emergence of classical spacetime [157, 156, 99, 182, 172, 113, 124, 158, 68, 8, 143, 81]. In particular, Ref. [158] showed that the areas of the extremal surfaces anchored to the boundaries of regions on a leaf of a holographic screen satisfy relations obeyed by entanglement entropies, so that they can indeed be identified as the entanglement entropies associated with the corresponding regions in the holographic space. We analyze properties of these surfaces and discuss their implications for a holographic theory of general spacetimes.

We lay down our general framework in Section 6.2. We then study the behavior of extremal surfaces in cosmological Friedmann-Robertson-Walker (FRW) spacetimes in Section 6.3. Here we focus on initially expanding flat and open universes, in which the area of the leaves monotonically increases. We first consider universes dominated by a single component in the Friedmann equation, and we identify how screen entanglement entropies—the entanglement entropies among the degrees of freedom in the holographic space—encode information about the spacetimes. We discuss next how the screen entanglement entropies behave in a transition period in which the dominant component of the universe changes. We find an interesting theorem when the holographic screen is spacelike: the change of a screen entanglement entropy is always monotonic. The proof of this theorem is given in Appendix 6.6. If the holographic screen is timelike, no such theorem holds.

In Section 6.4, we study the structure of the holographic theory for general spacetimes, building on the results obtained earlier. In particular, we discuss how the holographic entanglement entropies for general spacetimes differ from those in AdS/CFT and how, nevertheless, the former reduce to the latter in an appropriate limit. We emphasize that the holographic entanglement entropies for cosmological spacetimes obey a volume law, rather than an area law, implying that the relevant holographic states are not ground states of local field theories. This is the case despite the fact that the dynamics of the holographic theory respects some sense of locality, indicated by the fact that the area of a leaf increases in a local manner on a holographic screen.

The Hilbert space of the theory is analyzed in Section 6.4 under two assumptions:

(i) The holographic theory has (effectively) a qubit degree of freedom per each volume of $4\ln 2$ in Planck units. These degrees of freedom appear local at lengthscales larger than a microscopic cutoff $l_{\rm c}$.

(ii) If a holographic state represents a semiclassical spacetime, the area of an extremal surface anchored to the boundary of a region Γ on a leaf σ and contained in the causal region associated with σ represents the entanglement entropy of Γ in the holographic theory.

We find that these two assumptions strongly constrain the structure of the Hilbert space, although they do not determine it uniquely. There are essentially two possibilities:

Direct sum structure — Holographic states representing different semiclassical spacetimes $\mathcal{M}$ live in different Hilbert spaces $\mathcal{H}_{\mathcal{M}}$ even if these spacetimes have the same boundary space (or leaf) B

$$\mathcal{H}_B = \bigoplus_{\mathcal{M}} \mathcal{H}_{\mathcal{M}}. \tag{6.1}$$

In each Hilbert space $\mathcal{H}_{\mathcal{M}}$, the states representing the semiclassical spacetime comprise only a tiny subset of all the states—the vast majority of the states in $\mathcal{H}_{\mathcal{M}}$ do not allow for a semiclassical interpretation, which we call "firewall" states borrowing the terminology in Refs. [10, 9, 126]. In fact, the states allowing for a semiclassical spacetime interpretation do not even form a vector space—their superposition may lead to a firewall state if it involves a large number of terms, of order a positive power of $\dim \mathcal{H}_{\mathcal{M}}$. This is because a superposition involving such a large number of terms significantly alters the entanglement entropy structure, so under assumption (ii) above we cannot interpret the resulting state as a semiclassical state representing $\mathcal{M}$. In this picture, small excitations over spacetime $\mathcal{M}$ can be represented by standard linear operators acting on the (suitably extended) Hilbert space $\mathcal{H}_{\mathcal{M}}$, which can be trivially promoted to linear operators in $\mathcal{H}_B$.

Spacetime equals entanglement — Holographic states that represent different semiclassical spacetimes but have same boundary space B are all elements of a single Hilbert space $\mathcal{H}_B$. And yet, the number of independent microstates representing *each* of these spacetimes, $\mathcal{M}, \mathcal{M}', \mathcal{M}'', \cdots$, is the dimension of $\mathcal{H}_B$:

$$|\Psi_i^{\mathcal{M}}\rangle,\ |\Psi_{i'}^{\mathcal{M}'}\rangle,\ |\Psi_{i''}^{\mathcal{M}''}\rangle,\ \cdots \in \mathcal{H}_B; \qquad i, i', i'', \cdots = 1, \cdots, \dim \mathcal{H}_B, \tag{6.2}$$

which implies that the microstates representing different spacetimes are not independent. This picture arises if we require the converse of assumption (ii) and is called "spacetime equals entanglement" [143]: if a holographic state has the form of entanglement entropies corresponding to a certain spacetime, then the state indeed represents that spacetime. The structure of Eq. (6.2) is then obtained because arbitrary unitary transformations acting in each cutoff size cell in B do not change the entanglement entropies, implying that the number of microstates for any geometry is $\dim \mathcal{H}_B$ (so they span a basis of $\mathcal{H}_B$). Despite the intricate structure of the states, this picture admits the standard many worlds interpretation for classical spacetimes, as shown in Ref. [143]. Small excitations over spacetime are represented by non-linear/state-dependent operators, along the lines of Ref. [151] (see also [150, 183, 138]), since a superposition of background spacetimes may lead to another spacetime, so that operators representing excitations must know the entire quantum state they act on.

We note that a dichotomy similar to the one described above was discussed earlier in Ref. [151], but the interpretation and the context in which it appears here are distinct. First, the state-dependence of the operators representing excitations in the second scenario (as well as that of the time evolution operator) becomes relevant when the boundary space

is involved in the dynamics as in the case of cosmological spacetimes. Hence, this particular state-dependence need not persist in the AdS/CFT limit. This does not imply anything about the description of the interior a black hole in the CFT. It is possible that the CFT does not provide a semiclassical description of the black hole interior, i.e. it gives only a distant description. Alternatively, there may be a way of obtaining a state-dependent semiclassical description of the black hole interior within a CFT, as envisioned in Ref. [151]. We are agnostic about this issue.

Second, Ref. [151] describes the dichotomy as state-dependence vs. firewalls. Our picture, on the other hand, does not have a relation with firewalls because the following two statements apply to *both* the direct sum and spacetime equals entanglement pictures:

- Most of the states in the Hilbert space, e.g. in the Haar measure, are firewalls in the sense that they do not represent smooth semiclassical spacetimes, which require special entanglement structures among the holographic degrees of freedom.

- The fact that most of the states are firewalls does not mean that these states are realized as a result of standard time evolution, in which the volume of the boundary space increases in time. In fact, the direct sum picture even has a built-in mechanism of eliminating firewalls through time evolution, as we will see in Section 6.4.[1]

Rather, the real tension is between the linearity/state-independence of operators representing observables (including the time evolution operator) and the spacetime equals entanglement hypothesis, i.e. the hypothesis that if a holographic state has entanglement entropies corresponding to a semiclassical spacetime, then the state indeed represents that spacetime. If we insist on the linearity of observables, we are forced to take the direct sum picture; if we adopt the spacetime equals entanglement hypothesis, then we must give up linearity.

Our analysis in Section 6.4 also includes the following. In Section 6.4, we discuss bulk reconstruction from a holographic state, which suggests that the framework provides a distant description for a dynamical black hole. In Section 6.4, we consider how the theory encodes information about spacetime outside the causal region of a leaf, which is needed for autonomous time evolution. Our analysis suggests a strengthened covariant entropy bound: the entropy on the *union* of two light sheets (future-directed ingoing and past-directed outgoing) of a leaf is bounded by the area of the leaf divided by 4. This bound is stronger than the original bound in Ref. [27], which says that the entropy on *each* of the two light sheets is bounded by the area divided by 4. In Section 6.4, we analyze properties of time evolution, in particular a built-in mechanics of eliminating firewalls in the direct sum picture and the required non-linearity of the time evolution operator in the spacetime equals entanglement picture. In Sections 6.4 and 6.4, we discuss how our framework may reduce to AdS/CFT and string theory in an asymptotically Minkowski background in the appropriate limits. We

[1] This is natural because any dynamics leading to classicalization selects only a very special set of states as the result of time evolution: states interpreted as a superposition of a small number of classical worlds, where small means a number (exponentially) smaller than the dimension of the full microscopic Hilbert space.

argue that the dynamics of these theories (in which the boundaries are sent to infinity) describe that of the general holographic theory modded out by "vacuum degeneracies" relevant for the dynamics of the boundaries and the exteriors.

In Section 6.5, we devote our final discussion to the issue of selecting a state. In general, specifying a system requires selection conditions on a state in addition to determining the theory. To address this issue in quantum gravity, we need to study the problem of time [48, 188]. We discuss possible signals from a past singularity or past null infinity, closed universes and "fine-tuning" of states, and selection conditions for the string theory landscape [34, 104, 168, 53], especially the scenario called the "static quantum multiverse" [134]. While our discussion in this section is schematic, it allows us to develop intuition about how quantum gravity might work at the fundamental level when applied to the real world.

Throughout the paper, we adopt the Schrödinger picture of quantum mechanics and take the Planck length to be unity, $l_{\rm P} = 1$. When the semiclassical picture is applicable, we assume the null and causal energy conditions to be satisfied. These impose the conditions $\rho \geq -p$ and $|\rho| \geq |p|$, respectively, on the energy density ρ and pressure p of an ideal fluid component. The equation of state parameter $w = p/\rho$, therefore, takes a value in the range $|w| \leq 1$.

6.2 Holography and Quantum Gravity

The holographic principle states that quantum mechanics of a system with gravity can be formulated as a non-gravitational theory in spacetime with dimension one less than that in the gravitational description. The covariant entropy bound, or Bousso bound, [27] suggests that this holographically reduced—or "boundary"—spacetime may be identified as a hypersurface in the original gravitational spacetime determined by a collection of light rays. Specifically, it implies that the entropy on a null hypersurface generated by a congruence of light rays terminating at a caustic or singularity is bounded by its largest cross sectional area $\mathcal{A}$; in particular, the entropy on each side of the largest cross sectional surface is bounded by $\mathcal{A}/4$ in Planck units.[2] It is therefore natural to consider that, for a fixed gravitational spacetime, the holographic theory lives on a hypersurface—called the holographic screen—on which null hypersurfaces foliating the spacetime have the largest cross sectional areas [28].

This procedure of erecting a holographic screen has a large ambiguity, presumably reflecting a large freedom in fixing the redundancy of the gravitational description associated with the holographic principle. A particularly useful choice advocated in Refs. [132, 133, 140] is to adopt an "observer centric reference frame." Let the origin of the reference frame follow a timelike curve $p(\tau)$ which passes through a fixed spacetime point p_0 at $\tau = 0$, and consider the congruence of past-directed light rays emanating from p_0.[3] The expansion of

[2] We will conjecture a stronger bound in Section 6.4.

[3] In Refs. [132, 133, 140], $p(\tau)$ was chosen to be a timelike geodesic with τ being the proper time measured at $p(\tau)$. We suspect that this simplifies the time evolution operator in the holographic theory.

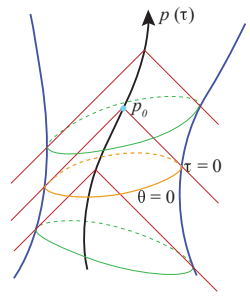


Figure 6.1: For a fixed semiclassical spacetime, the holographic screen is a hypersurface obtained as the collection of codimension-2 surfaces (labeled by τ) on which the expansion of the light rays emanating from a timelike curve $p(\tau)$ vanishes, $\theta = 0$. This way of erecting the holographic screen automatically deals with the redundancy associated with complementarity. The ambiguity of choosing $p(\tau)$ reflects a large freedom in fixing the redundancy associated with holography.

the light rays θ satisfies

$$\frac{\partial\theta}{\partial\lambda} + \frac{1}{2}\theta^2 \leq 0, \tag{6.3}$$

where λ is the affine parameter associated with the light rays. This implies that the light rays emitted from p_0 focus toward the past (starting from $\theta = +\infty$ at $\lambda = 0_+$), and we may identify the apparent horizon, i.e. the codimension-2 surface with

$$\theta = 0, \tag{6.4}$$

to be an equal-time hypersurface—called a leaf—of a holographic screen. Repeating the procedure for all τ, we obtain a specific holographic screen, with the leaves parameterized by τ, corresponding to foliating the spacetime region accessible to the observer at $p(\tau)$; see Fig. 6.1. Such a foliation is consonant with the complementarity hypothesis [170], which asserts that a complete description of a system is obtained by referring only to the spacetime region that can be accessed by a single observer.

With this construction, we can view a quantum state of the holographic theory as living on a leaf of the holographic screen obtained in the above observer centric manner. We can then consider the collection of all possible quantum states on all possible leaves, obtained by considering all timelike curves in all spacetimes. We take the view that a state of quantum gravity lives in the Hilbert space spanned by all of these states (together with other states that do not admit a full spacetime interpretation) [132, 133]. It is often convenient to

consider a Hilbert space $\mathcal{H}_B$ spanned by the holographic states that live on the "same" boundary space B.[4] The relevant Hilbert space can then be written as

$$\mathcal{H} = \sum_B \mathcal{H}_B, \tag{6.5}$$

where the sum of Hilbert spaces is defined by[5]

$$\mathcal{H}_1 + \mathcal{H}_2 = \{v_1 + v_2 \,|\, v_1 \in \mathcal{H}_1, v_2 \in \mathcal{H}_2\}. \tag{6.6}$$

This formulation is not restricted to descriptions based on fixed semiclassical spacetime backgrounds. For example, we may consider a state in which macroscopically different spacetimes are superposed; in particular, this picture describes the eternally inflating multiverse as a state in which macroscopically different universes are superposed [132, 134]. The space in Eq. (6.5) is called the covariant Hilbert space with observer centric gauge fixing.

Recently, Bousso and Engelhardt have identified two special classes of holographic screens [31, 32]: if a portion of a holographic screen is foliated by marginally anti-trapped (trapped) surfaces, then that portion is called a past (future) holographic screen. Specifically, denoting the two future-directed null vector fields orthogonal to a portion of a leaf by k^a and l^a, with k^a being tangent to light rays emanating from $p(\tau)$, the expansion of the null geodesic congruence generated by l^a satisfies $\theta_l > 0$ and < 0 for past and future holographic screens, respectively. They proved, building on earlier works [89, 90, 14, 13], that the area of leaves $\mathcal{A}(\tau)$ monotonically increases (decreases) for a past (future) holographic screen:

$$\begin{cases} \theta_k = 0 \\ \theta_l \gtrless 0 \end{cases} \quad \Leftrightarrow \quad \frac{d}{d\tau}\mathcal{A}(\tau) \gtrless 0; \tag{6.7}$$

see Fig. 6.2. In many regular circumstances, including expanding FRW universes, the holographic screen is a past holographic screen, so that the area of the leaves monotonically increases, $d\mathcal{A}(\tau)/d\tau > 0$. In this paper we mostly focus on this case, and we interpret the area theorem in terms of the second law of thermodynamics applied to the Hilbert space of Eq. (6.5). Moreover, in Ref. [159] it was proved that this area theorem holds locally on the holographic screen: the area of any fixed spatial portion of the holographic screen, determined by a vector field tangent to the holographic screen and normal to its leaves, increases

[4]The exact way in which the boundary spaces are grouped into different B's is unimportant. For example, one can regard the boundary spaces having the same area $\mathcal{A}$ within some precision $\delta\mathcal{A}$ to be in the same B, or one can discriminate them further by their induced metrics. This ambiguity does not affect any of the results, unless one takes $\delta\mathcal{A}$ to be exponentially small in $\mathcal{A}$ or discriminates induced metrics with the accuracy of order the Planck length (which corresponds to resolving microstates of the spacetime).

[5]Unlike Ref. [133], here we do not assume specific relations between $\mathcal{H}_B$'s; for example, $\mathcal{H}_{B_1}$ and $\mathcal{H}_{B_2}$ for different boundary spaces B_1 and B_2 may not be orthogonal. Also, we have included in the sum over B the cases in which B is outside the semiclassical regime, i.e. the cases in which the holographic space does not correspond to a leaf of a holographic screen in a semiclassical regime. These issues will be discussed in Section 6.4.

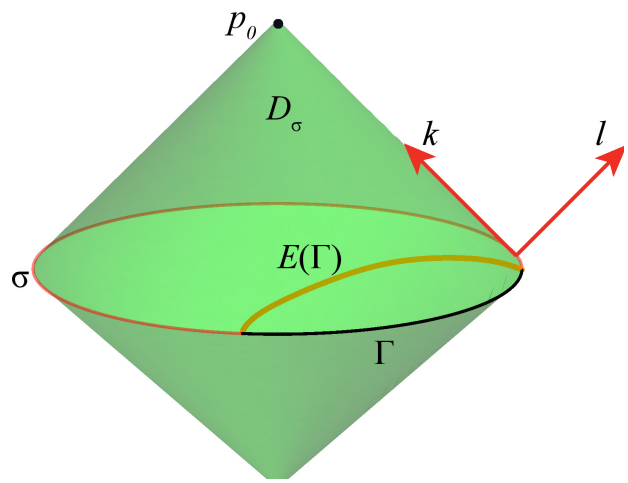


Figure 6.2: The congruence of past-directed light rays emanating from p_0 (the origin of the reference frame) has the largest cross sectional area on a leaf σ, where the holographic theory lives. At any point on σ, there are two future-directed null vectors orthogonal to the leaf: k^a and l^a. For a given region Γ of the leaf, we can find a codimension-2 extremal surface $E(\Gamma)$ anchored to the boundary $\partial\Gamma$ of Γ, which is fully contained in the causal region D_σ associated with σ.

monotonically in time. This implies that the dynamics of the holographic theory respects some notion of locality.

What is the structure of the holographic theory and how can we explore it? Recently, a conjecture has been made in Ref. [158] which relates geometries of general spacetimes in the gravitational description to the entanglement entropies of states in the holographic theory. This extends the analogous theorem/conjecture in the AdS/CFT context [157, 156, 99] to more general cases, allowing us to probe the holographic description of general spacetimes, including those that do not have an obvious spacetime boundary on which the holographic theory can live. In particular, Ref. [158] proved that for a given region Γ of a leaf σ, a codimension-2 extremal surface $E(\Gamma)$ anchored to the boundary $\partial\Gamma$ of Γ is fully contained in the causal region D_σ of σ:

$$D_\sigma : \text{ the domain of dependence of an interior achronal hypersurface whose only boundary is } \sigma, \tag{6.8}$$

where the concept of the interior is defined so that a vector on σ pointing toward the interior takes the form $c_1 k^a - c_2 l^a$ with $c_1, c_2 > 0$ (see Fig. 6.2). This implies that the normalized area of the extremal surface $E(\Gamma)$

$$S(\Gamma) = \frac{1}{4}\|E(\Gamma)\|, \tag{6.9}$$

satisfies expected properties of entanglement entropy, such as strong subadditivity, so that

it can be identified with the entanglement entropy of the region Γ in the holographic theory. Here, $\|x\|$ represents the area of x. If there are multiple extremal surfaces in D_σ for a given Γ, then we must take the one with the minimal area.

In the rest of the paper, we study the holographic theory of quantum gravity for general spacetimes, adopting the framework described in this section. We first analyze FRW spacetimes and then discuss lessons learned from that analysis later.

6.3 Holographic Description of FRW Universes

In this section, we study the putative holographic description of $(3+1)$-dimensional FRW cosmological spacetimes:

$$ds^2 = -dt^2 + a^2(t)\left[\frac{dr^2}{1-\kappa r^2} + r^2(d\psi^2 + \sin^2\psi\, d\phi^2)\right], \tag{6.10}$$

where $a(t)$ is the scale factor, and $\kappa < 0$, $= 0$ and > 0 for open, flat and closed universes, respectively. The Friedmann equation is given by

$$\left(\frac{\dot{a}}{a}\right)^2 + \frac{\kappa}{a^2} = \frac{8\pi}{3}\rho, \tag{6.11}$$

where the dot represents t derivative. Here, we include the energy density from the cosmological constant as a component in ρ having the equation of state parameter $w = -1$.

As discussed in the previous section, we describe the system as viewed from a reference frame whose origin follows a timelike curve $p(\tau)$, which we choose to be at $r = 0$. The holographic theory then lives on the holographic screen, an equal-time slice of which is an apparent horizon: a codimension-2 surface on which the expansion of the light rays emanating from $p(\tau)$ for a fixed τ vanishes. Under generic conditions, this horizon is always at a finite distance

$$r = \frac{1}{\sqrt{\dot{a}^2(t_*) + \kappa}} \equiv r_{\rm AH}(t_*) < \infty, \tag{6.12}$$

where t_* is the FRW time on the horizon. Note that the symmetry of the setup makes the FRW time the same everywhere on the apparent horizon, and for an open universe, $\dot{a}(t_*) > \sqrt{-\kappa}$ is satisfied for values of τ before $p(\tau)$ hits the big crunch. For flat and open universes, we find that this surface is always marginally anti-trapped, i.e. a leaf of a past holographic screen, as long as the universe is initially expanding. On the other hand, for a closed universe the surface can change from marginally anti-trapped to marginally trapped as τ increases, implying that the holographic screen may be a past holographic screen only until certain time τ. In this section, we focus our attention on initially expanding flat and open universes. Closed universes will be discussed in Section 6.5.

Below, we study entanglement entropies for subregions in the holographic theory—screen entanglement entropies—adopting the conjecture of Ref. [158]. Here we focus on studying the

properties of these entropies, leaving their detailed interpretation for later. We first discuss "stationary" aspects of screen entanglement entropies, concentrating on states representing spacetime in which the expansion of the universe is dominated by a single component in the Friedmann equation. We study how screen entanglement entropies encode the information about the spacetime the state represents. We then analyze dynamics of screen entanglement entropies during a transition period in which the dominant component changes. Implications of these results in the broader context of the holographic description of quantum gravity will be discussed in the next section.

Holographic dictionary for FRW universes

Consider a Hilbert space $\mathcal{H}_B$ spanned by a set of quantum states living in the same codimension-2 boundary surface B. As mentioned in footnote 4, the definition of the boundary surface being the same has an ambiguity. For our analysis of states representing FRW spacetimes, we take the boundary B to be specified by its area $\mathcal{A}_B$ (within some precision $\delta\mathcal{A}_B$ that is not exponentially small in $\mathcal{A}_B$). In this subsection, we focus on a single Hilbert space $\mathcal{H}_* \in \{\mathcal{H}_B\}$ specified by a fixed (though arbitrary) boundary area $\mathcal{A}_*$.

Consider FRW universes with $\kappa \leq 0$ having vacuum energy ρ_Λ and filled with varying ideal fluid components.[6] For every universe with

$$\rho_\Lambda < \frac{3}{2\mathcal{A}_*}, \tag{6.13}$$

there is an FRW time t_* at which the area of the leaf of the past holographic screen is $\mathcal{A}_*$; see Fig. 6.3. This is because the area of the leaf of the past holographic screen is monotonically increasing [31], and the final (asymptotic) value of the area is given by

$$\mathcal{A}_\infty = \begin{cases} \frac{3}{2\rho_\Lambda}, & \text{for } \rho_\Lambda > 0, \\ +\infty, & \text{for } \rho_\Lambda \leq 0. \end{cases} \tag{6.14}$$

Any quantum state representing the system at any such moment is an element of $\mathcal{H}_*$. A question is what features of the holographic state encode information about the universe it represents.

To study this problem, we perform the following analysis. First, given an FRW universe specified by the history of the energy density of the universe, $\rho(t)$, we determine the FRW time t_* at which the apparent horizon σ_*, identified as a leaf of the past holographic screen, has the area $\mathcal{A}_*$:

$$\begin{cases} \rho(t) \\ \mathcal{A}_* \end{cases} \rightarrow \; t_*, \tag{6.15}$$

[6] The ρ_Λ here represents the energy density of a (local) minimum of the potential near which fields in the FRW universe in question take values. In fact, string theory suggests that there is no absolutely stable de Sitter vacuum in full quantum gravity; it must decay, at least, before the Poincaré recurrence time [104].

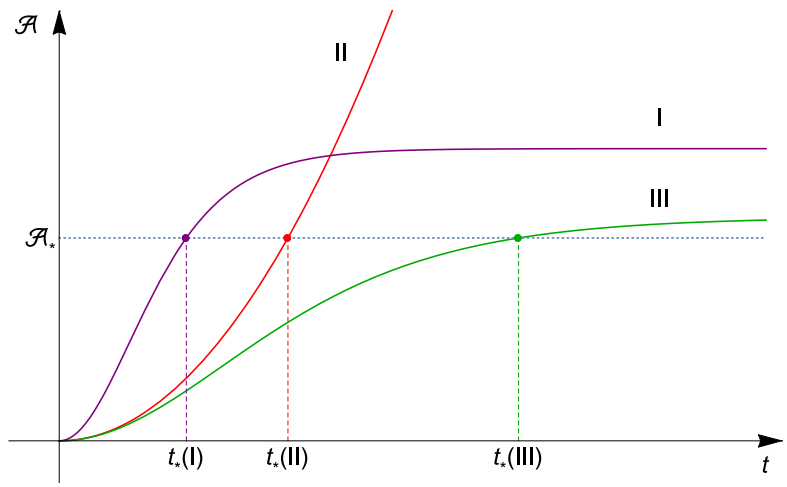


Figure 6.3: Various FRW universes, $\mathrm{I}, \mathrm{II}, \mathrm{III}, \cdots$, have the same boundary area $\mathcal{A}_*$ at different times, $t_*(\mathrm{I}), t_*(\mathrm{II}), t_*(\mathrm{III}), \cdots$. Quantum states representing universes at these moments belong to Hilbert space $\mathcal{H}_*$ specified by the value of the boundary area.

where we assume Eq. (6.13). We then consider a spherical cap region of the leaf σ_* specified by an angle γ ($0 \leq \gamma \leq \pi$):

$$L(\gamma): \quad t = t_*, \quad r = r_{\mathrm{AH}}(t_*), \quad 0 \leq \psi \leq \gamma, \tag{6.16}$$

where $r_{\mathrm{AH}}(t_*)$ is given by Eq. (6.12) (see Fig. 6.4), and determine the extremal surface $E(\gamma)$ which is codimension-2 in spacetime, anchored on the boundary of $L(\gamma)$, and fully contained inside the causal region D_{σ_*} associated with σ_*. According to Ref. [158], we interpret the quantity

$$S(\gamma) = \frac{1}{4}\|E(\gamma)\|, \tag{6.17}$$

to represent von Neumann entropy of the holographic state representing the region $L(\gamma)$, obtained after tracing out the complementary region on σ_*.

To determine the extremal surface $E(\gamma)$, it is useful to introduce cylindrical coordinates

$$\xi = r \sin\psi, \qquad z = r\cos\psi. \tag{6.18}$$

We find that the isometry of the FRW metric, Eq. (6.10), allows us to move the boundary on which the extremal surface is anchored, $\partial L(\gamma)$, on the $z = 0$ plane:

$$\partial L(\gamma): \quad t = t_*, \quad \xi = r_{\mathrm{AH}}(t_*)\sin\gamma \equiv \xi_{\mathrm{AH}}, \quad z = 0. \tag{6.19}$$

The surface to be extremized is then parameterized by functions $t(\xi)$ and $z(\xi)$ with the boundary conditions

$$t(\xi_{\mathrm{AH}}) = t_*, \qquad z(\xi_{\mathrm{AH}}) = 0, \tag{6.20}$$

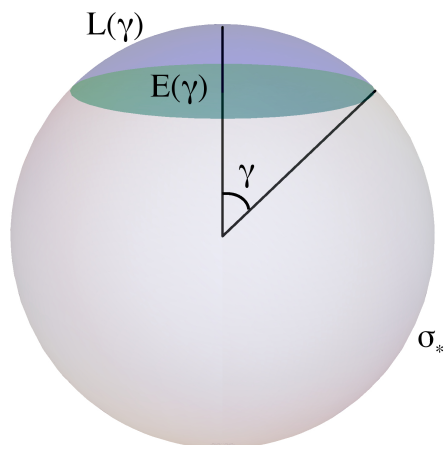


Figure 6.4: A region $L(\gamma)$ of the leaf σ_* is parameterized by an angle $\gamma : [0, \pi]$. The extremal surface $E(\gamma)$ anchored to its boundary, $\partial L(\gamma)$, is also depicted schematically. (In fact, $E(\gamma)$ bulges into the time direction.)

and the area functional to be extremized is given by

$$2\pi \int_0^{\xi_{\rm AH}} a(t)\, \xi \sqrt{-\left(\frac{dt}{d\xi}\right)^2 + \frac{a^2(t)}{1-\kappa(\xi^2+z^2)}\left\{(1-\kappa z^2) + (1-\kappa\xi^2)\left(\frac{dz}{d\xi}\right)^2 + 2\kappa\xi z \frac{dz}{d\xi}\right\}}\, d\xi. \tag{6.21}$$

In all the examples we study (in this and next subsections), we find that the extremal surface does not bulge into the z direction. In this case, we can set $z = 0$ in Eq. (6.21) and find

$$\|E(\gamma)\| = \underset{t(\xi)}{\rm ext} \left[2\pi \int_0^{r_{\rm AH}(t_*)\sin\gamma} a(t)\, \xi \sqrt{-\left(\frac{dt}{d\xi}\right)^2 + \frac{a^2(t)}{1-\kappa\xi^2}}\, d\xi \right]. \tag{6.22}$$

The analysis described above is greatly simplified if the expansion of the universe is determined by a single component in the Friedmann equation, i.e. a single fluid component with the equation of state parameter w or negative spacetime curvature. We thus focus on the case in which the expansion is dominated by a single component in (most of) the region probed by the extremal surfaces. In realistic FRW universes this holds for almost all t, except for a few Hubble times around when the dominant component changes from one to another. Discussion about a transition period in which the dominant component changes will be given in the next subsection.

A flat FRW universe filled with a single fluid component

Suppose the expansion of the universe is determined dominantly by a single ideal fluid component with w. The scale factor is then given by

$$a(t) = c\, t^{\frac{2}{3(1+w)}}, \tag{6.23}$$

where c is a constant, and the metric in the region $r \leq r_{\rm AH}$ takes the form

$$ds^2 = -dt^2 + c^2\, t^{\frac{4}{3(1+w)}} \left[dr^2 + r^2(d\psi^2 + \sin^2\psi\, d\phi^2)\right], \tag{6.24}$$

where we have used the fact that $|\kappa\, r_{\rm AH}^2| \ll 1$. In this case, we find that the $\mathcal{A}_*$ dependence of screen entanglement entropy S_Γ for an arbitrarily shaped region Γ on σ_*—specified as a region on the ψ-ϕ plane—is given by

$$S_\Gamma = \tilde{S}_\Gamma \mathcal{A}_*, \tag{6.25}$$

where $\tilde{S}_\Gamma$ does not depend on $\mathcal{A}_*$. This can be seen in the following way.

Consider the causal region D_{σ_*} associated with σ_*. For certain values of w (i.e. $w \geq 1/3$), D_{σ_*} hits the big bang singularity. It is thus more convenient to discuss the "upper half" of the region:

$$D^+_{\sigma_*} = \{p \in D_{\sigma_*} \mid t(p) \geq t_*\}. \tag{6.26}$$

In an expanding universe, the extremal surface anchored on the boundary of a region Γ on σ_* is fully contained in this region. Now, by performing t_*-dependent coordinate transformation

$$\rho = \frac{2}{3(1+w)} c\, t_*^{-\frac{1+3w}{3(1+w)}} r, \tag{6.27}$$

$$\eta = \frac{2}{3(1+w)} \left[\left(\frac{t}{t_*}\right)^{\frac{1+3w}{3(1+w)}} - 1\right], \tag{6.28}$$

the region $D^+_{\sigma_*}$ is mapped into

$$0 \leq \eta \leq 1, \qquad 0 \leq \rho \leq 1 - \eta, \tag{6.29}$$

and the metric in $D^+_{\sigma_*}$ is given by

$$ds^2\big|_{D^+_{\sigma_*}} = \frac{\mathcal{A}_*}{4\pi} \left(\frac{1+3w}{2}\eta + 1\right)^{\frac{4}{1+3w}} \left[-d\eta^2 + d\rho^2 + \rho^2(d\psi^2 + \sin^2\psi\, d\phi^2)\right], \tag{6.30}$$

where

$$\mathcal{A}_* = 9\pi(1+w)^2 t_*^2. \tag{6.31}$$

Since $\mathcal{A}_*$ appears only as an overall factor of the metric in Eqs. (6.29, 6.30), we conclude that the $\mathcal{A}_*$ dependence of $S_\Gamma \propto \|E_\Gamma\|$ is only through an overall proportionality factor, as in Eq. (6.25).

Due to the scaling in Eq. (6.25), it is useful to consider an object obtained by dividing S_Γ by a quantity that is also proportional to $\mathcal{A}_*$. We find it convenient to define the quantity

$$Q_\Gamma \equiv \frac{S_\Gamma}{V_\Gamma/4}, \tag{6.32}$$

where V_Γ is the (2-dimensional) "volume" of the region Γ or its complement $\bar{\Gamma}$ on the boundary surface σ_*, whichever is smaller. This quantity is independent of $\mathcal{A}_*$, and hence t_*. For the spherical region of Eq. (6.16), we find

$$Q(\gamma) = \frac{S(\gamma)}{V(\gamma)/4} = \frac{\|E(\gamma)\|}{V(\gamma)}, \tag{6.33}$$

where

$$V(\gamma) = \frac{1}{2}\Big\{1 - \mathrm{sgn}\Big(\frac{\pi}{2} - \gamma\Big)\cos\gamma\Big\}\mathcal{A}_*. \tag{6.34}$$

An explicit expression for $Q(\gamma)$ is given by

$$Q(\gamma) = \frac{1}{1 - \mathrm{sgn}(\frac{\pi}{2} - \gamma)\cos\gamma} \underset{f(x)}{\mathrm{ext}} \left[\int_0^{\sin\gamma} x\, f^{\frac{4}{1+3w}} \sqrt{1 - \Big(\frac{2}{1+3w}\Big)^2 \Big(\frac{df}{dx}\Big)^2}\, dx\right], \tag{6.35}$$

where the extremization with respect to function $f(x)$ is performed with the boundary condition

$$f(\sin\gamma) = 1, \tag{6.36}$$

and we have used the fact that the extremal surface does not bulge into the z direction in the cylindrical coordinates of Eq. (6.18). From the point of view of the holographic theory, Q_Γ represents the amount of entanglement entropy per degree of freedom as viewed from the smaller of Γ and $\bar{\Gamma}$. As we will discuss in Section 6.4, the fact that this is a physically significant quantity has important implications for the structure of the holographic theory.

In Fig. 6.5, we plot $Q(\gamma)$ as a function of γ ($0 \le \gamma \le \pi/2$) for various values of w: -1 (vacuum energy), -0.98, -0.8, 0 (matter), 1/3 (radiation), and 1. The value of $Q(\gamma)$ for $\pi/2 \le \gamma \le \pi$ is given by $Q(\gamma) = Q(\pi - \gamma)$. We find the following features:

- In the limit of a small boundary region, $\gamma \ll 1$, the value of $Q(\gamma)$ approaches unity regardless of the value of w:

  $$Q_w(\gamma) \xrightarrow{\gamma \ll 1} 1. \tag{6.37}$$

  This implies that for a small boundary region, the entanglement entropy of the region is given by its volume in the holographic theory in Planck units:

  $$S_w(\gamma) \xrightarrow{\gamma \ll 1} \frac{1}{4} V(\gamma). \tag{6.38}$$

  For larger γ ($\le \pi/2$), $Q(\gamma)$ becomes monotonically small as γ increases:

  $$\frac{d}{d\gamma} Q_w(\gamma) < 0. \tag{6.39}$$

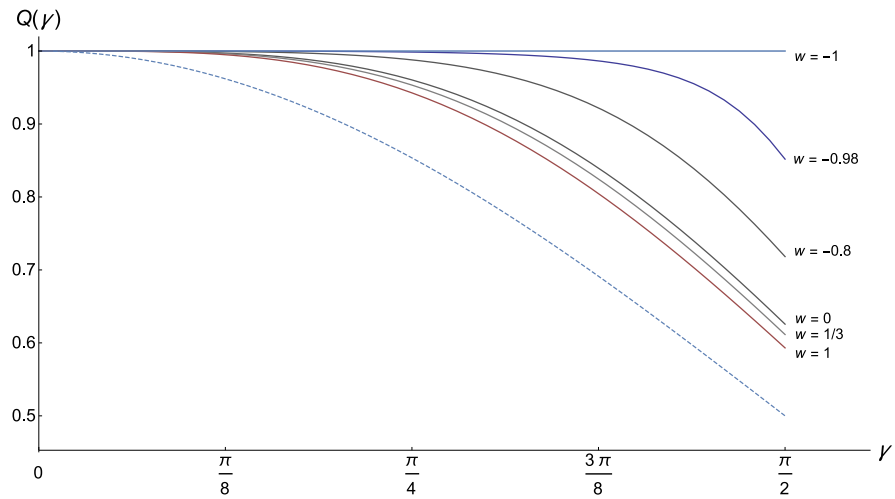


Figure 6.5: The value of $Q(\gamma)$ as a function of γ ($0 \leq \gamma \leq \pi/2$) for $w = -1$ (vacuum energy), -0.98, -0.8, 0 (matter), 1/3 (radiation), and 1. The dotted line indicates the lower bound given by the flat space geometry, which can be realized in a curvature dominated open FRW universe.

The deviation of $Q(\gamma)$ from 1 near $\gamma = 0$ is given by

$$Q_w(\gamma) \overset{\gamma \ll 1}{=} 1 - c\,(1+w)\gamma^4 + \cdots\,, \tag{6.40}$$

where $c > 0$ is a constant that does not depend on w.

- For any fixed boundary region, γ, the value of $Q(\gamma)$ decreases monotonically in w:

$$\frac{d}{dw} Q_w(\gamma) < 0. \tag{6.41}$$

In particular, when w approaches -1 (from above), $Q(\gamma)$ becomes unity:

$$\lim_{w \to -1} Q_w(\gamma) = 1. \tag{6.42}$$

This implies that in the limit of de Sitter FRW ($w \to -1$), the state in the holographic theory becomes "randomly entangled" (i.e. saturates the Page curve [144]):[7]

$$\lim_{w \to -1} S_w(\gamma) = \frac{1}{4} V(\gamma). \tag{6.43}$$

[7] In the case of an exactly single component with $w = -1$, the expansion of light rays emanating from p_0, i.e. θ_k, becomes 0 only at infinite affine parameter λ. We view this as a result of mathematical idealization. A realistic de Sitter FRW universe is obtained by introducing an infinitesimally small amount of matter in addition to the $w = -1$ component, which avoids the above issue. The results obtained in this way agree with those by first taking $w > -1$ and then the limit $w \to -1$.

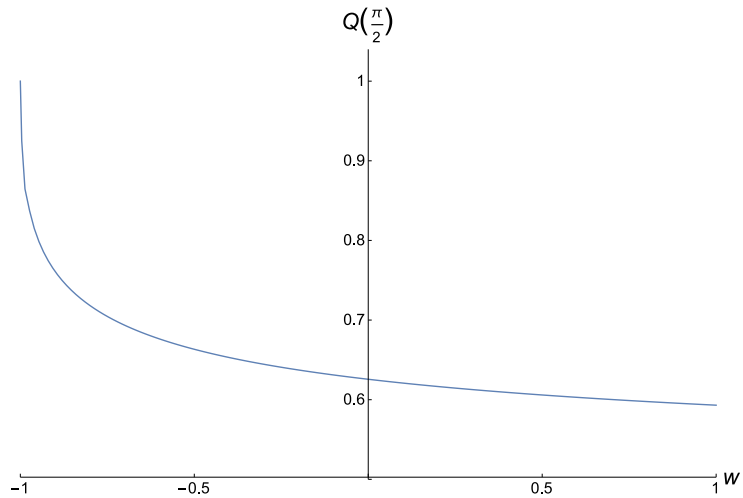


Figure 6.6: The value of $Q(\pi/2)$ as a function of w.

Note that $V(\gamma)$ is the smaller of the volume of $L(\gamma)$ and that of its complement on the leaf. The value of $Q(\pi/2)$ (the case in which $L(\gamma)$ is a half of the leaf) is plotted as a function of w in Fig. 6.6.

We will discuss further implications of these findings in Section 6.4.

We note that there are simple geometric bounds on the values of $Q_w(\gamma)$. This can be seen by adopting the maximin construction [158, 187]: the extremal surface is the one having the maximal area among all possible codimension-2 surfaces each of which is anchored on $\partial L(\gamma)$ and has minimal area on some interior achronal hypersurface bounded by σ. This implies that the area of the extremal surface, $\|E(\gamma)\|$, cannot be larger than the boundary volume $V(\gamma)$, giving $Q(\gamma) \leq 1$. Also, the extremal surface cannot have a smaller area than the codimension-2 surface that has the minimal area on a constant time hypersurface $t = t_*$: $\|E(\gamma)\| \geq \pi\{a(t_*)r_{\rm AH}(t_*)\sin\gamma\}^2$. Together, we obtain

$$\frac{\sin^2\gamma}{2\left\{1 - \mathrm{sgn}(\frac{\pi}{2} - \gamma)\cos\gamma\right\}} \leq Q_w(\gamma) \leq 1. \tag{6.44}$$

The lower edge of this range is depicted by the dashed line in Fig. 6.5. We find that the upper bound of Eq. (6.44) can be saturated with $w \to -1$, while the lower bound cannot with $|w| \leq 1$. If we formally take $w \to +\infty$, the lower bound can be reached. A fluid with $w > 1$, however, does not satisfy the causal energy condition (although it satisfies the null energy condition), so we do not consider such a component.

As a final remark, we show in Fig. 6.7 the shape of the extremal surface for $\gamma = \pi/2$ for the same values of w as in Fig. 6.5: -1, -0.98, -0.8, 0, 1/3, and 1. The horizontal axis is the cylindrical radial coordinate normalized by the apparent horizon radius, $\xi/\xi_{\rm AH}$, and the

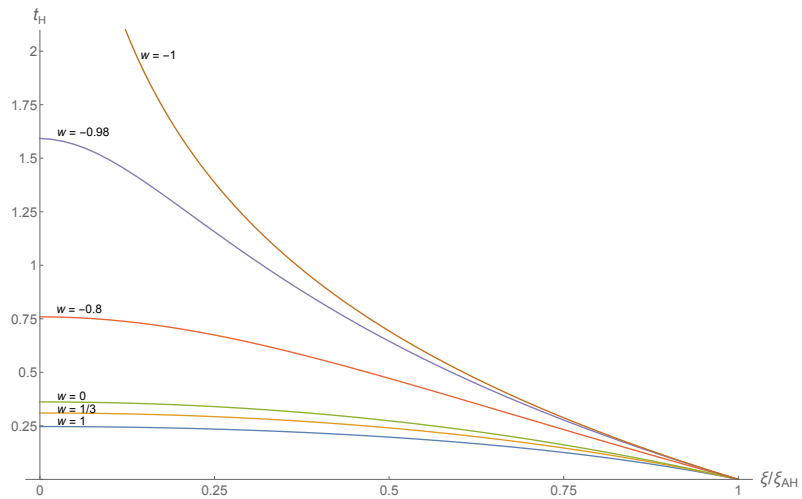


Figure 6.7: The shape of the extremal surfaces $E(\pi/2)$ for $w = -1$, -0.98, -0.8, 0, 1/3, and 1. The horizontal axis is the cylindrical radial coordinate normalized by the apparent horizon radius, $\xi/\xi_{\rm AH}$, and the vertical axis is the Hubble time, $t_{\rm H}$.

vertical axis is taken to be the Hubble time defined by

$$t_{\rm H} = \int_{t_*}^{t} \frac{\dot{a}(t)}{a(t)}\, dt = \frac{2}{3(1+w)} \ln \frac{t}{t_*}, \tag{6.45}$$

which reduces in the $w \to -1$ limit to the usual Hubble time $t_{\rm H} = H(t - t_*)$, where $H = \dot{a}/a$. We find that the extremal surface bulges into the future direction for any w. In fact, this occurs generally in an expanding universe and can be understood from the maximin construction: the scale factor increases toward the future, so that the area of the minimal area surface on an achronal hypersurface increases when the hypersurface bulges into the future direction in time. The amount of the bulge is $t_{\rm H} \approx O(1)$, except when $w \approx -1$. For $w \to -1$, the extremal surface probes $t_H \to +\infty$ as $\xi/\xi_{\rm AH} \to +0$, but its area is still finite, $\|E(\pi/2)\| \to \mathcal{A}_*/2$, as the surface becomes almost null in this limit.

An open FRW universe dominated by curvature

We now consider an open FRW universe dominated by curvature, i.e. the case in which the expansion of the universe is determined by the second term in the left-hand side of Eq. (6.11). This implies that the distance to the apparent horizon is much larger than the curvature length scale

$$\frac{-\kappa}{a^2(t)} \gg \frac{8\pi}{3}\rho(t) \quad \Longleftrightarrow \quad r_{\rm AH}(t) \gg \frac{1}{\sqrt{-\kappa}} \equiv r_{\rm curv}. \tag{6.46}$$

(Note that $\kappa < 0$ for an open universe.) As seen in Eqs. (6.11, 6.12), the value of $r_{\rm AH}(t)$ is determined by $\rho(t)$, which gives only a minor contribution to the expansion of the universe.

The scale factor is given by

$$a(t) = \sqrt{-\kappa}\, t. \tag{6.47}$$

The extremal surface can be found easily by noticing that the universe in this limit is a hyperbolic foliation of a portion of the Minkowski space: the coordinate transformation

$$\tilde{t} = t\sqrt{1 + \left(\sqrt{-\kappa}\, r\right)^2}, \tag{6.48}$$

$$\tilde{r} = \sqrt{-\kappa}\, t\, r, \tag{6.49}$$

leads to the Minkowski metric $ds^2 = -d\tilde{t}^2 + d\tilde{r}^2 + \tilde{r}^2(d\psi^2 + \sin^2\psi\, d\phi^2)$. The extremal surface is thus a plane on a constant $\tilde{t}$ hypersurface, which in the FRW (cylindrical) coordinates is given by

$$t_{\rm H} \approx \ln \frac{1}{\xi/\xi_{\rm AH}} \qquad (0 \le \xi/\xi_{\rm AH} \le 1), \tag{6.50}$$

where $\xi_{\rm AH} = r_{\rm AH}(t_*)\sin\gamma$, and $t_{\rm H}$ is the Hubble time

$$t_{\rm H} = \int_{t_*}^{t} \frac{\dot{a}(t)}{a(t)}\, dt = \ln \frac{t}{t_*}. \tag{6.51}$$

The resulting $Q(\gamma)$ is

$$Q(\gamma) \approx \frac{\sin^2\gamma}{2\left\{1 - {\rm sgn}(\frac{\pi}{2} - \gamma)\cos\gamma\right\}}. \tag{6.52}$$

This, in fact, saturates the lower bound in Eq. (6.44), plotted as the dashed line in Fig. 6.5.

Dynamics of screen entanglement entropies in a transition

Let us consider the evolution of an FRW universe. From the holographic theory point of view, it is described by a time-dependent state $|\Psi(\tau)\rangle$ living on $\sigma(\tau)$. Because of the area theorem of Refs. [31, 32], we can take τ to be a monotonic function of the leaf area, leading to

$$\frac{d}{d\tau}\mathcal{A}(\tau) > 0, \tag{6.53}$$

where $\mathcal{A}(\tau) \equiv \|\sigma(\tau)\|$. This evolution involves a change in the number of (effective) degrees of freedom, $\mathcal{A}(\tau)/4$, as well as that of the structure of entanglement on the boundary, $Q_\Gamma(\tau)$. For the latter, we mostly consider $Q(\gamma,\tau)$ associated with a spherical cap region $\Gamma = L(\gamma)$. A natural question is if a statement similar to Eq. (6.53) applies for screen entanglement entropies:

$$\frac{d}{d\tau}S(\gamma,\tau) \overset{?}{>} 0. \tag{6.54}$$

Here,

$$S(\gamma,\tau) = Q(\gamma,\tau)\frac{V(\gamma,\tau)}{4}, \tag{6.55}$$

with

$$V(\gamma,\tau) = \frac{1}{2}\Big\{1 - \mathrm{sgn}\Big(\frac{\pi}{2} - \gamma\Big)\cos\gamma\Big\}\mathcal{A}(\tau), \tag{6.56}$$

being the smaller of the boundary volumes of $L(\gamma)$ and its complement.

There are some cases in which we can show that the relation in Eq. (6.54) is indeed satisfied. Consider, for example, a flat FRW universe filled with various fluid components having differing equations of states: w_i $(i = 1, 2, \cdots)$. As time passes, the dominant component of the universe changes from one having larger w to one having smaller w successively. This implies that $Q(\gamma,\tau)$ monotonically increases in time, so that Eq. (6.53) indeed implies Eq. (6.54) in this case. Another interesting case is when the holographic screen is spacelike. In this case, we can prove that the time dependence of $S(\gamma,\tau)$ is monotonic; see Appendix 6.6. In particular, if we have a spacelike past holographic screen (which occurs for $w > 1/3$ in a single-component dominated flat FRW universe), then the screen entanglement entropy for an arbitrary region increases in time: $dS_\Gamma(\tau)/d\tau > 0$.

What happens if the holographic screen is timelike? One might think that there is an obvious argument against the inequality in Eq. (6.54). Suppose the expansion of the early universe is dominated by a fluid component with w. Suppose at some FRW time t_0 this component is converted into another fluid component having a different equation of state parameter w', e.g. by reheating associated with the decay of a scalar condensate. If $w' > w$, then the Q value after the transition is smaller than that before

$$Q_{w'}(\gamma) - Q_w(\gamma) < 0. \tag{6.57}$$

One may think that this can easily overpower the increase of $S(\gamma,\tau)$ from the increase of the area: $d\mathcal{A}(\gamma,\tau)/d\tau > 0$ [159]. In particular, if w is close to -1, then the increase of the area before the transition is very slow, so that the effect of Eq. (6.57) would win over that of the area increase. However, as depicted in Fig. 6.7, when $w \approx -1$ the extremal surface bulges into larger t by many Hubble times. Hence the time between the moments in which Eq. (6.55) can be used before and after the transition becomes long, opening the possibility that the relevant area increase is non-negligible.

To make the above discussion more explicit, let us compare the values of the screen entanglement entropy $S(\gamma)$ corresponding to two extremal surfaces depicted in Fig. 6.8: the "latest" extremal surface that is fully contained in the w region and the "earliest" extremal surface fully contained in the w' region, each anchored to the leaves at FRW times t_* and t_0. This provides the most stringent test of the inequality in Eq. (6.54) that can be performed using the expression of Eq. (6.55) for fixed w's. The ratio of the entanglement entropies is given by

$$R_{w'w}(\gamma) \equiv \frac{S_{\rm after}(\gamma)}{S_{\rm before}(\gamma)} = \frac{Q_{w'}(\gamma)}{Q_w(\gamma)}\frac{t_0^2}{t_*^2} = \frac{Q_{w'}(\gamma)}{Q_w(\gamma)}\, e^{3(1+w)t_{{\rm H},w}}, \tag{6.58}$$

where $t_{{\rm H},w}$ is the Hubble time between t_* and t_0, given by Eq. (6.45) with $t \to t_0$. In Fig. 6.9, we plot $R_w \equiv R_{1w}(\pi/2)$; setting $w' = 1$ minimizes the ratio. We find that this ratio can

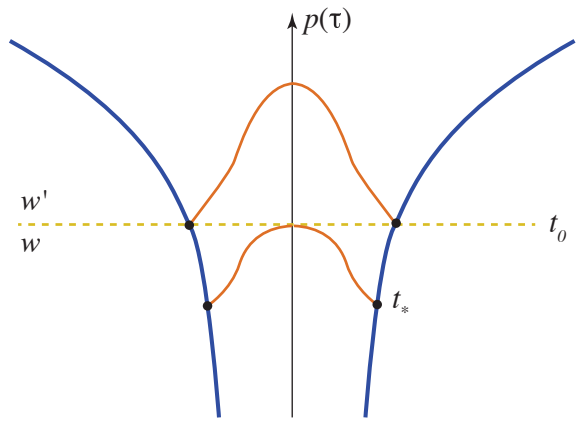


Figure 6.8: An FRW universe whose dominant component changes from w to w' at time t_0. Two surfaces depicted by orange lines are the latest extremal surface fully contained in the w region (bottom) and the earliest extremal surface fully contained in the w' region (top), each anchored to the leaves at t_* and t_0.

be smaller than 1 for $w \approx -1$. In fact, for $w \to -1$ we find the value obtained naively by assuming that the area does not change before the transition:

$$R_{-1} = \frac{Q_1\left(\frac{\pi}{2}\right)}{Q_{-1}\left(\frac{\pi}{2}\right)} = Q_1\left(\frac{\pi}{2}\right), \tag{6.59}$$

although for $w = -1$ there is no such thing as the latest extremal surface that is fully contained in the region before the transition (since $t_{\mathrm{H},-1} = +\infty$).

This analysis suggests that screen entanglement entropies can in fact drop if the system experiences a rapid transition induced by some dynamics,[8] although the instantaneous transition approximation adopted above is not fully realistic. Of course, such a drop is expected to be only a temporary phenomenon—because of the area increase after the transition, the entropy generally returns back to the value before the transition in a characteristic dynamical timescale and then continues to increase afterward. We expect that the relation in Eq. (6.54) is valid in a coarse-grained sense

$$\frac{d}{d\tau}\bar{S}(\gamma,\tau) > 0; \qquad \bar{S}(\gamma,\tau) = \frac{1}{\tau_c}\int_{\tau}^{\tau+\tau_c} S(\gamma,\tau')\,d\tau', \tag{6.60}$$

but not "microscopically" in general. Here, τ_c must be taken sufficiently larger than the characteristic dynamical timescale, the Hubble time for an FRW universe.

[8]This does not mean that the second law of thermodynamics is violated. The entropy discussed here is the von Neumann entropy of a significant portion (half) of the whole system, which can deviate from the thermodynamic entropy of the region when the system experiences a rapid change.

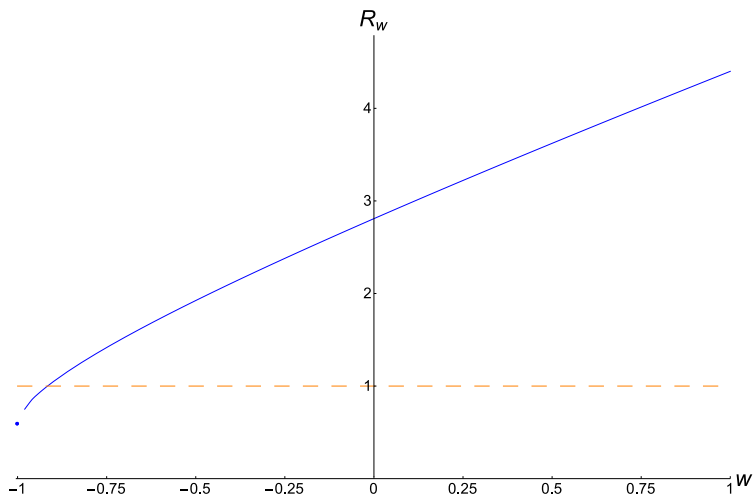


Figure 6.9: The ratio of the screen entanglement entropies, $R_w = R_{1w}(\pi/2)$, before and after the transition from a universe with the equation of state parameter w to that with $w' = 1$, obtained from Figs. 6.6 and 6.7 using Eq. (6.58). The dot at $w = -1$ represents $R_{-1} = R_{1-1}(\pi/2)$ obtained in Eq. (6.59).

For further illustration, we perform numerical calculations for how the area of a leaf hemisphere, $\|L(\pi/2, t)\|$, and the associated screen entanglement entropy, calculated using $S(\pi/2, t) = \|E(\pi/2, t)\|/4$, evolve in time during transitions from a $w = -1$ to a $w' = 0$ flat FRW universe. Here, we take the FRW time t as the time parameter. For this purpose, we consider a scalar field ϕ having a potential $V(\phi)$ that has a flat portion and a well, with the initial value of ϕ being in the flat portion. We first note that a transformation of the potential of the form

$$V(\phi) \to V'(\phi) = \epsilon^2 V(\phi), \tag{6.61}$$

leads to rescalings of the scalar field, $\phi(t)$, and the scale factor, $a(t)$, obtained as the solutions to the equations of motion:

$$\phi'(t) = \phi(\epsilon t), \qquad a'(t) = a(\epsilon t). \tag{6.62}$$

Plugging these in Eq. (6.22), we find that the area functionals before and after the transformation Eq. (6.61) are related by simple rescaling $t \to t/\epsilon$ and $\xi \to \xi/\epsilon$, so that

$$\left\|E'\left(\frac{\pi}{2}, t\right)\right\| = \frac{1}{\epsilon^2}\left\|E\left(\frac{\pi}{2}, \frac{t}{\epsilon}\right)\right\|. \tag{6.63}$$

These scaling properties imply that the leaf hemisphere area and the screen entanglement entropy for the transformed potential are read off from those for the untransformed one by

$$\left\|L'\left(\frac{\pi}{2}, t\right)\right\| = \frac{1}{\epsilon^2}\left\|L\left(\frac{\pi}{2}, \frac{t}{\epsilon}\right)\right\|, \qquad \left\|S'\left(\frac{\pi}{2}, t\right)\right\| = \frac{1}{\epsilon^2}\left\|S\left(\frac{\pi}{2}, \frac{t}{\epsilon}\right)\right\|. \tag{6.64}$$

We therefore need to be concerned only with the shape of the potential, not its overall scale. In particular, we can always be in the semiclassical regime by performing a transformation with $\epsilon \ll 1$.

In Fig. 6.10, we show the results of our calculations for "steep" and "broad" potentials. The explicit forms of the potentials are given by

$$V(\phi) = 1 - e^{-k(\phi-\phi_0)^2} + s(\phi - \phi_0)\tanh(p(\phi - \phi_0)), \tag{6.65}$$

with

$$\text{Steep}: \quad k = 5000, \quad s = 0.01, \quad p = 20, \quad \phi_0 = 0.045, \tag{6.66}$$

$$\text{Broad}: \quad k = 25, \quad s = 0.01, \quad p = 2, \quad \phi_0 = 0.5, \tag{6.67}$$

although their detailed forms are unimportant. For the steep potential, plotted in Fig. 6.10(a), we show the time evolutions of $\phi(t)$, $\|L(\pi/2, t)\|$, and $S(\pi/2, t)$ in Figs. 6.10(b)–(d) for the initial conditions of $\phi(0) = \dot{\phi}(0) = 0$ and $a(0) = 0.01$. The same are shown for the broad potential, Fig. 6.10(e), in Figs. 6.10(f)–(h) for the initial conditions of $\phi(0) = \dot{\phi}(0) = 0$ and $a(0) = 10^{-11}$. In either cases, the leaf hemisphere area increases monotonically while the screen entanglement entropy experiences drops as the field oscillates around the minimum. The fractional drops from the first, second, and third peaks are $\simeq 1.3\%$, 0.9%, and 0.6%, respectively, for the steep potential and $\simeq 2.5\%$, 1.6%, and 1.2%, respectively, for the broad potential.

We thus find that screen entanglement entropies may decrease in a transition period. The interpretation of this result, however, needs care. Since the system is far from being in a "vacuum" during a transition, true entanglement entropies for subregions in the holographic theory may have contributions beyond that captured by the simple formula of Eq. (6.17). This would require corrections of the formula, possibly along the lines of Refs. [64, 63, 30], and with such corrections the drop of the entanglement entropy we have found here might disappear. We leave a detailed study of this issue to future work.

6.4 Interpretation: Beyond AdS/CFT

The entanglement entropies in the holographic theory of FRW universes seen so far show features different from those in CFTs of the AdS/CFT correspondence. Here we highlight these differences and see how properties characteristic to local CFTs are reproduced when bulk spacetime becomes asymptotically AdS. We also discuss implications of our findings for the structure of the holographic theory. In particular, we discuss the structure of the Hilbert space for quantum gravity applicable to general spacetimes. While we cannot determine the structure uniquely, we can classify possibilities under certain general assumptions. The issues discussed include bulk reconstruction, the interior and exterior regions of the leaf, and time evolution in the holographic theory.

Volume/area law for screen entanglement entropies

One can immediately see that holographic entanglement entropies for FRW universes have two features that are distinct from those in AdS/CFT. First, unlike entanglement entropies in CFTs, the holographic entanglement entropies for FRW universes are finite for a finite value of $\mathcal{A}_*$. Second, as seen in Section 6.3, e.g. Eq. (6.25), these entropies obey a volume law, rather than an area law.[9] (Note that $\mathcal{A}_*$ is a volume from the viewpoint of the holographic theory.) In particular, in the limit that the region Γ in the holographic theory becomes small, the entanglement entropy S_Γ becomes proportional to the volume V_Γ *with a universal coefficient*, which we identified as 1/4 to match the conventional results in Refs. [21, 22, 20, 84, 86, 70]. (For a small enough subsystem, we expect that the entanglement entropy agrees with the thermal entropy.) From the bulk point of view, this is because the extremal surface E_Γ approaches Γ itself, so that $\|E_\Gamma\| \to V_\Gamma$.

What do these features mean for the holographic theory? The finiteness of the entanglement entropies implies that the cutoff length of the holographic theory is finite, i.e. the number of degrees of freedom in the holographic theory is finite, at least effectively. In particular, our identification implies that the holographic theory effectively has a qubit degree of freedom per volume of $4\ln 2$ (in Planck units), although it does not mean that the cutoff length of the theory is necessarily $\simeq \sqrt{4\ln 2}$. It is possible that the cutoff length is $l_{\rm c} > \sqrt{4\ln 2}$ and that each cutoff size cell has $N = l_{\rm c}^2/4\ln 2$ (> 1) degrees of freedom. In fact, since the string length $l_{\rm s}$ and the Planck length are related as $l_{\rm s}^2 \sim n$, where n is the number of species in the low energy theory (including the moduli fields parameterizing the extra dimensions) [55], it seems natural to identify $l_{\rm c}$ and N as $l_{\rm s}$ and n, respectively.

The volume law of the entangled entropies implies that a holographic state corresponding to an FRW universe is not a ground state of local field theory, which is known to satisfy an area law [26, 163]. This does not necessarily mean that the holographic theory for FRW universes must be nonlocal at lengthscales larger than the cutoff $l_{\rm c}$; it might simply be that the relevant states are highly excited ones. In fact, the dynamics of the holographic theory is expected to respect some aspects of locality as suggested by the fact that the area theorem applies locally on a holographic screen [159]. Of course, it is also possible that the holographic states for FRW universes are states of some special class of nonlocal theories.

The features of screen entangled entropies described here are not specific to FRW universes but appear in more general "cosmological" spacetimes, spacetimes in which the holographic screen is at finite distances and the gravitational dynamics is not frozen there. If the interior region of the holographic screen is (asymptotically) AdS, these features change. In this case, the same procedure as in Section 6.2 puts the holographic screen at spatial infinity (the AdS boundary), and the AdS geometry makes the area of the extremal surface anchored to the boundary $\partial\Gamma$ of a small region Γ on a leaf proportional to the area of $\partial\Gamma$ with a diverging coefficient: $\|E_\Gamma\| \sim \|\partial\Gamma\|/\epsilon$ ($\epsilon \to 0$). This makes the screen entanglement entropies obey an area law, so that the holographic theory can now be a ground state of a

[9] A similar property was argued for holographic entropies for Euclidean flat spacetime in Ref. [114].

local field theory. In fact, the theory is a CFT [122, 74, 189], consistent with the fact that we could take the cutoff length to zero, $l_c \sim \epsilon \to 0$.

The structure of holographic Hilbert space

We now discuss implications of our analysis for the structure of the Hilbert space of quantum gravity for general spacetimes. We work in the framework of Section 6.2; in particular, we assume that when a holographic state represents a semiclassical spacetime, the area of the extremal surface contained in D_σ and anchored to the boundary of a region Γ on a leaf represents the entanglement entropy of the region Γ in the holographic theory, Eq. (6.9). Note that this does not necessarily mean that the converse is true; there may be a holographic state in which entanglement entropies for subregions do not correspond to the areas of extremal surfaces in a semiclassical spacetime.

Consider a holographic state representing an FRW spacetime. The fact that for a small enough region Γ the area of the extremal surface anchored to its boundary approaches the volume of the region on the leaf, $\|E_\Gamma\| \to V_\Gamma$, implies that the degrees of freedom in the holographic theory are localized and that their density is, at least effectively, one qubit per $4\ln 2$ (although the cutoff length of the theory may be larger than $\sqrt{4\ln 2}$). We take these for granted as anticipated in the original holographic picture [173, 169]. This suggests that the number of holographic degrees of freedom which comprise FRW states on the leaf σ_* with area $\mathcal{A}_*$ is $\mathcal{A}_*/4$ *for any value of* w.

Given these assumptions, there are still a few possibilities for the structure of the Hilbert space of the holographic theory. Below we enumerate these possibilities and discuss their salient features.

Direct sum structure

Let us first assume that state vectors representing FRW universes with different w's are independent of each other, as indicated in the left portion of Fig. 6.11. This implies that the Hilbert space $\mathcal{H}_* \in \{\mathcal{H}_B\}$, which contains holographic states for FRW universes at times when the leaf area is $\mathcal{A}_*$, has a direct sum structure

$$\mathcal{H}_* = \bigoplus_w \mathcal{H}_{*,w}. \tag{6.68}$$

Here, we regard universes with the equation of state parameters falling in a range $\delta w \ll 1$ to be macroscopically identical, where δw is a small number that does not scale with $\mathcal{A}_*$.[10] This is the structure envisioned originally in Ref. [133].

What is the structure of $\mathcal{H}_{*,w}$? A natural possibility is that each of these subspaces has dimension

$$\ln \dim \mathcal{H}_{*,w} = \frac{\mathcal{A}_*}{4}. \tag{6.69}$$

[10] If we consider FRW universes with multiple fluid components, the corresponding spaces must be added in the right-hand side of Eq. (6.68).

This is motivated by the fact that arbitrary unitary transformations acting in each cutoff size cell do not change the structure of screen entanglement entropies, and they can lead to $e^{\mathcal{A}_*/4}$ independent holographic states that have the screen entanglement entropies corresponding to the FRW universe with the equation of state parameter w. If we regard all of these states as microstates for the FRW universe with w, then we obtain Eq. (6.69). This, however, does not mean that the holographic states representing the FRW universe with w comprise the Hilbert space $\mathcal{H}_{*,w}$. Since these states form a basis of $\mathcal{H}_{*,w}$, their superposition can lead to a state which has entanglement entropies far from those corresponding to the FRW universe with w. In fact, we can even form a state in which degrees of freedom in different cells are not entangled at all. This is a manifestation of the fact that entanglement cannot be represented by a linear operator.

This implies that states representing the semiclassical FRW universe are "preferred basis states" in $\mathcal{H}_{*,w}$, and their arbitrary linear combinations may lead to states that do not admit a semiclassical interpretation. We expect that these preferred axes are "fat": we have to superpose a large number of basis states, in fact exponentially many in $\mathcal{A}_*$, to obtain a state that is not semiclassical (because we need that many states to appreciably change the entanglement structure, as illustrated in a toy qubit model in Appendix 6.6). It is, however, true that most of the states in $\mathcal{H}_{*,w}$, including those having the entanglement entropy structure corresponding to a universe with another w, are states that do not admit a semiclassical spacetime interpretation. Drawing an analogy with the work in Refs. [10, 9, 126], we may call them "firewall" states. In Section 6.4, we argue that these states are unlikely to be produced by standard semiclassical time evolution.

The dimension of $\mathcal{H}_*$ is given by

$$\ln \dim \mathcal{H}_* = \ln \sum_w e^{\frac{\mathcal{A}_*}{4}} \approx \frac{\mathcal{A}_*}{4} - \ln \delta w \simeq \frac{\mathcal{A}_*}{4}, \tag{6.70}$$

as expected from the covariant entropy bound (unless δw is exponentially small in $\mathcal{A}_*$, which we assume not to be the case). Small excitations over the FRW universes may be represented in suitably extended spaces $\mathcal{H}_{*,w}$. Since entropies associated with the excitations are typically subdominant in $\mathcal{A}_*$ [173, 142], they have only minor effects on the overall picture, e.g. Eq. (6.70). (Note that the excitations here do not contain the degrees of freedom attributed to gravitational, e.g. Gibbons-Hawking, radiation. These degrees of freedom are identified as the microscopic degrees of freedom of spacetimes, i.e. the vacuum degrees of freedom [136, 137, 135], which are already included in Eq. (6.69).) The operators representing the excitations can be standard linear operators acting on the Hilbert space $\mathcal{H}_*$, at least in principle.

We also mention the possibility that the logarithm of the number of independent states N_w representing the FRW universe with w is smaller than $\mathcal{A}_*/4$. For example, it might be given approximately by twice the entanglement entropy for a leaf hemisphere $S_w(\pi/2) = Q_w(\pi/2)\mathcal{A}_*/8$:

$$\ln N_w \approx Q_w\left(\frac{\pi}{2}\right)\frac{\mathcal{A}_*}{4}. \tag{6.71}$$

The basic picture in this case is not much different from that discussed above; for example, the difference of the values of $\ln \dim \mathcal{H}_*$ is higher order in $1/\mathcal{A}_*$ (although this possibility makes the issue of the equivalence condition for the boundary space label B nontrivial). We will not consider this case separately below.

Russian doll structure: spacetime equals entanglement

In the picture described above, the structures of $\mathcal{H}_{*,w}$'s are all very similar. Each of these spaces has the dimension of $\mathcal{A}_*/4$ and has $e^{\mathcal{A}_*/4}$ independent states that represent the FRW universe with a fixed value of w. An arbitrary linear combination of these states, however, is not necessarily a state representing the FRW universe with w. In the previous picture, we identified all such states as the firewall (or unphysical) states, but is it possible that some of these states, in fact, represent other FRW universes? In particular, is it possible that all the $\mathcal{H}_{*,w}$ spaces are actually the *same* space, i.e. $\mathcal{H}_{*,w_1} = \mathcal{H}_{*,w_2}$ for all $w_1 \neq w_2$?

A motivation to consider this possibility comes from the fact that if w does not by itself provide an independent label for states, then the $e^{\mathcal{A}_*/4}$ independent microstates for the FRW universe with a fixed w can form a basis for the configuration space of the $\mathcal{A}_*/4$ holographic degrees of freedom. This implies that we can superpose these states to obtain many—in fact $e^{\mathcal{A}_*/4}$—independent states that have the entanglement entropies corresponding to the FRW universe with any $w' \neq w$, which we can identify as the states representing the FRW universe with w'.[11] In essence, this amounts to saying that the converse of the statement made at the beginning of this subsection is true: when a holographic state has the form of entanglement entropies corresponding to a certain spacetime, then the state indeed represents that spacetime. This scenario was proposed in Ref. [143] and called "spacetime equals entanglement." It is depicted in the right portion of Fig. 6.11.

One might think that the scenario does not make sense, since it implies that a superposition of classical universes can lead to a different classical universe. Wouldn't it make any reasonable many worlds interpretation of spacetime impossible? In Ref. [143], it was argued that this is not the case. First, for a given FRW universe, we expect that the space of its microstates is "fat"; namely, a superposition of less than $e^{O(\delta w \mathcal{A}_*)}$ microstates representing a classical universe leads only to another microstate representing the same universe. This implies that the $e^{\mathcal{A}_*/4}$ microstates of a classical universe generate an "effective vector space," unless we consider a superposition of an exponentially large, $\gtrsim e^{O(\delta w \mathcal{A}_*)}$, number of states.

What about a superposition of different classical universes? In particular, if states representing universes with w_1 and w_2 ($\neq w_1$) are superposed, then how does the theory know that the resulting state represents a superposition of two classical universes, and not another—perhaps even non-classical—universe? A key point is that the Hilbert space we consider has

[11] The same argument applies to the FRW universes with multiple fluid components, so that the states representing these universes also live in the same Hilbert space as the single component universes.

a special basis, determined by the $\mathcal{A}_*/4$ local degrees of freedom in the holographic space:[12]

$$\mathcal{H}_* = (\mathbf{C}^2)^{\otimes \frac{\mathcal{A}_*}{4}}. \tag{6.72}$$

From the result in Section 6.3, we know that a state representing the FRW universe with w_1 is more entangled than that representing the FRW universe with w_2 ($> w_1$). This implies that when expanded in the natural basis $\{|\Psi_i\rangle\}$ for the structure of Eq. (6.72), i.e. the product state basis for the $\mathcal{A}_*/4$ local holographic degrees of freedom, then a state $|\Psi_{w_1}\rangle$ representing the universe with w_1 effectively has exponentially more terms than a state $|\Psi_{w_2}\rangle$ representing the universe with w_2. Namely, we expect that

$$|\Psi_w\rangle \approx \sum_{i=1}^{e^{f(w)\frac{\mathcal{A}_*}{4}}} a_i\,|\Psi_i\rangle, \tag{6.73}$$

where $f(w)$ is a monotonically decreasing function of w taking values of $O(1)$, and a_i are coefficients taking generic random values. The normalization condition for $|\Psi_w\rangle$ then implies

$$|a_i| \approx O\big(e^{-f(w)\frac{\mathcal{A}_*}{8}}\big), \tag{6.74}$$

i.e. the size of the coefficients in product basis expansion is exponentially different for states with different w's. This, in particular, leads to

$$\langle\Psi_{w_1}|\Psi_{w_2}\rangle \lesssim O\big(e^{-\{f(w_1)-f(w_2)\}\frac{\mathcal{A}_*}{8}}\big), \tag{6.75}$$

i.e. microstates for different universes are orthogonal up to exponentially suppressed corrections.

Now consider a superposition state

$$|\Psi\rangle = c_1|\Psi_{w_1}\rangle + c_2|\Psi_{w_2}\rangle, \tag{6.76}$$

where $|c_1|^2 + |c_2|^2 = 1$ up to the correction from exponentially small overlap $\langle\Psi_{w_1}|\Psi_{w_2}\rangle$. We are interested in the reduced density matrix for a subregion Γ in the holographic theory

$$\rho_\Gamma = \mathrm{Tr}_{\bar{\Gamma}}\,|\Psi\rangle\langle\Psi|, \tag{6.77}$$

where Γ occupies less than a half of the leaf volume. The property of Eq. (6.75) then ensures that

$$\rho_\Gamma = |c_1|^2\rho_\Gamma^{(1)} + |c_2|^2\rho_\Gamma^{(2)}, \tag{6.78}$$

up to corrections exponentially suppressed in $\mathcal{A}_*$. Here, $\rho_\Gamma^{(1)}$ ($\rho_\Gamma^{(2)}$) are the reduced density matrices we would obtain if the state were genuinely $|\Psi_{w_1}\rangle$ ($|\Psi_{w_2}\rangle$). The matrix ρ_Γ thus takes

[12] For simplicity, here we have assumed that the degrees of freedom are qubits, but the subsequent argument persists as long as the number of independent states for each degree of freedom does not scale with $\mathcal{A}_*$. In particular, it persists if the correct structure of $\mathcal{H}_*$ appears as $(\mathbf{C}^N)^{\otimes \mathcal{A}_*/l_c^2}$ as discussed in Section 6.4.

the form of an incoherent classical mixture for the two universes. Similarly, the entanglement entropy for the region Γ is also incoherently added

$$S_\Gamma = |c_1|^2 S_\Gamma^{(1)} + |c_2|^2 S_\Gamma^{(2)} + S_{\Gamma,\mathrm{mix}}, \tag{6.79}$$

where $S_\Gamma^{(1,2)}$ are the entanglement entropies obtained if the state were $|\Psi_{w_{1,2}}\rangle$, and

$$S_{\Gamma,\mathrm{mix}} = -|c_1|^2 \ln |c_1|^2 - |c_2|^2 \ln |c_2|^2, \tag{6.80}$$

is the entropy of mixing (classical Shannon entropy), suppressed by factors of $O(\mathcal{A}_*)$ compared with $S_\Gamma^{(1,2)}$. The features in Eqs. (6.78, 6.79) indicate that unless $|c_1|$ or $|c_2|$ is suppressed exponentially in $\mathcal{A}_*$, the state $|\Psi\rangle$ admits the usual interpretation of a superposition of macroscopically different universes with $w_{1,2}$.

In fact, unless a superposition involves exponentially many microstates, we find

$$|\Psi\rangle = \sum_i c_i |\Psi_{w_i}\rangle \quad \Rightarrow \quad \begin{array}{l} \rho_\Gamma = \sum_i |c_i|^2 \rho_\Gamma^{(i)}, \\ S_\Gamma = \sum_i |c_i|^2 S_\Gamma^{(i)} + S_{\Gamma,\mathrm{mix}}, \end{array} \tag{6.81}$$

with exponential accuracy. Here, $S_{\Gamma,\mathrm{mix}} = -\sum_i |c_i|^2 \ln |c_i|^2$ and is suppressed by a factor of $O(\mathcal{A}_*)$ compared with the first term in S_Γ. This indicates that the standard many worlds interpretation applies to classical spacetimes under any reasonable measurements (only) in the limit that $e^{-\mathcal{A}_*}$ is regarded as zero, i.e. unless a superposition involves exponentially many terms or an exponentially small coefficient. This is consonant with the observation that classical spacetime has an intrinsically thermodynamic nature [100], supporting the idea that it consists of a large number of degrees of freedom. In Ref. [143], the features described above were discussed using a qubit model in which the states representing the FRW universes exhibit a "Russian doll" structure as illustrated in Fig 6.11. We summarize this model in Appendix 6.6 for completeness.

We conclude that the states representing FRW universes with a leaf area $\mathcal{A}_*$ can all be elements of a single Hilbert space $\mathcal{H}_*$ with dimension

$$\ln \dim \mathcal{H}_* = \frac{\mathcal{A}_*}{4}. \tag{6.82}$$

Any such universe has $e^{\mathcal{A}_*/4}$ independent microstates, which form a basis of $\mathcal{H}_*$. This implies that matter and spacetime must have a sort of unified origin in this picture, since a superposition that changes the spacetime geometry must also change the matter content filling the universe. How could this be the case?

Consider, as discussed in Section 6.4, that the cutoff length of the holographic theory is of order $l_\mathrm{s} \sim \sqrt{n}$, where n (> 1) is the number of species at energies below $1/l_\mathrm{s}$. This implies that the $\mathcal{A}_*/4$ degrees of freedom can be decomposed as

$$\frac{\mathcal{A}_*}{4} \sim n \frac{\mathcal{A}_*}{l_\mathrm{s}^2}, \tag{6.83}$$

representing n fields living in the holographic space of cutoff length l_s. Now, to obtain the $e^{\mathcal{A}_*/4}$ microstates for an FRW universe we need to consider rotations for all the n degrees of freedom in each cutoff size cell. This may suggest that the identity of a matter species at the fundamental level may not be as adamant as in low energy semiclassical field theories. The reason why all the n degrees of freedom can be involved could be because the "local effective temperature," defined analogously to de Sitter space, diverges at the holographic screen.

Finally, we expect that small excitations over FRW universes in this picture are represented by non-linear/state-dependent operators in the (suitably extended) $\mathcal{H}_*$ space, along the lines of Ref. [151] (see Refs. [150, 183, 138] for earlier work). This is because a superposition of background spacetimes may lead to another background spacetime, so that operators representing excitations should know the entire quantum state they are acting on.

Bulk reconstruction from holographic states

We have seen that the entanglement entropies of the $\mathcal{A}_*/4$ local holographic degrees of freedom in the holographic space σ_* encode information about spacetime in the causal region D_{σ_*}. Here we discuss in more detail how this encoding may work in general.

While we have focused on the case in which the future-directed ingoing light rays emanating orthogonally from σ_* (i.e. in the k^a directions in Fig. 6.2) meet at a point p_0, our discussion can be naturally extended to the case in which the light rays encounter a spacetime singularity before reaching a caustic. This may occur, for example, if a black hole forms in a universe as depicted in Fig. 6.12, where we have assumed spherical symmetry for collapsing matter and taken $p(\tau)$ to follow its center. We see that at intermediate times, the future-directed ingoing light rays emanating from leaves encounter the black hole singularity before reaching a caustic.[13] Our interpretation in this case is similar to the case without a singularity. The entanglement entropies of the holographic degrees of freedom encode information about D_{σ_*}.

In what sense does a holographic state on σ_* contain information about D_{σ_*}? We assume that the theory allows for the reconstruction of D_{σ_*} from the data in the state on σ_*. On the other hand, it is not the case that the collection of extremal surfaces for all possible subregions on σ_* probes the entire D_{σ_*}. This suggests that the full reconstruction of D_{σ_*} may need bulk time evolution.

There is, however, no a priori guarantee that the operation corresponding to bulk time evolution is complete within $\mathcal{H}_*$. This means that there may be no arrangement of operators defined in $\mathcal{H}_*$ that represents certain operators in D_{σ_*}. For these subsets of D_{σ_*}, bulk reconstruction would involve operators defined on other boundary spaces. In other words, the operators supported purely in $\mathcal{H}_*$ may allow for a direct spacetime interpretation only for a portion of D_{σ_*}, e.g. the outside of the black hole horizon in the example of Fig. 6.12 (in

[13] At these times, the specific construction of the holographic screen in Section 6.2 cannot be applied exactly. This is not a problem as the fundamental object is the state in the holographic space, and not $p(\tau)$. The purpose of the discussion in Section 6.2 is to illustrate our observer centric choice of fixing the holographic redundancy in formulating the holographic theory.

which case some of the operators would represent the stretched horizon degrees of freedom). Our assumption merely says that the operators in $\mathcal{H}_*$ acting on the state contain data equivalent to specifying the system on a Cauchy surface for D_{σ_*}.

The consideration above implies that the information in a holographic state on σ_*, when interpreted through operators in $\mathcal{H}_*$, may only be partly semiclassical. We expect that this becomes important when the spacetime has a horizon. In particular, for the $w = -1$ FRW universe, the leaf σ_* is formally beyond the stretched de Sitter horizon as viewed from $p(\tau)$. This may mean that some of the degrees of freedom represented by operators defined in $\mathcal{H}_*$ can only be viewed as non-semiclassical stretched horizon degrees of freedom.

Information about the "exterior" region

The information about D_{σ_*}, contained in the screen entanglement entropies, is not sufficient to determine future states obtained by time evolution. This information corresponds to that on the "interior" light sheet, i.e. the light sheet generated by light rays emanating in the $+k^a$ directions from σ_*.[14] However, even barring the possibility of information sent into the system from a past singularity or past null infinity (which we will discuss in Section 6.5), determining a future state requires information about the "exterior" light sheet, i.e. the one generated by light rays emanating in the $-k^a$ directions; see Fig. 6.13.[15] How is this information encoded in the holographic state? Does it require additional holographic degrees of freedom beyond the $\mathcal{A}_*/4$ degrees of freedom considered so far?

The simplest possibility is that the $e^{\mathcal{A}_*/4}$ microstates for each interior geometry (i.e. a fixed screen entanglement entropy structure) contain all the information associated with both the interior and exterior light sheets. If this is indeed the case, then we do not need any other degrees of freedom in the holographic space σ_* beyond the $\mathcal{A}_*/4$ ones discussed earlier. It also implies the following properties for the holographic theory:

- **Autonomous time evolution** — Assuming the absence of a signal sent in from a past singularity or past null infinity (see Section 6.5), the evolution of the state is autonomous. In particular, an initial pure state evolves into a pure state.

- ***S*-matrix description for a dynamical black hole** — As a special case, a pure state representing initial collapsing matter to form a black hole will evolve into a pure state representing final Hawking radiation, even if $p(\tau)$ hits the singularity at an intermediate stage (at least if the leaf stays outside the black hole); see Fig. 6.12.

- **Strengthened covariant entropy bound** — According to the original proposal of the covariant entropy bound [27, 29], the entropy on *each* of the interior and exterior

[14] If the light sheet encounters a singularity before reaching a caustic, then the information about the singularity may also be contained.

[15] This light sheet is terminated at a singularity or a caustic. Note that the information beyond a caustic is not needed to specify the state [140], since it is timelike related with the information on the interior light sheet [186] so that the two do not provide independent information.

light sheets is bounded by $\mathcal{A}_*/4$, implying that

$$\ln \dim \mathcal{H}_* = 2 \times \frac{\mathcal{A}_*}{4} = \frac{\mathcal{A}_*}{2}, \tag{6.84}$$

where $\mathcal{H}_*$ is the Hilbert space associated with σ_*. The present picture instead says

$$\ln \dim \mathcal{H}_* = \frac{\mathcal{A}_*}{4}, \tag{6.85}$$

implying that the entropy on the *union* of the interior and exterior light sheets is bounded by $\mathcal{A}_*/4$:[16] Note that the bound does not say that the entropy on each of the interior and exterior light sheets is separately bounded by $\mathcal{A}_*/8$, and so is profoundly holographic. This bound is consistent with the fact that in any known realistic examples the covariant entropy bound is saturated only in one side of a leaf [33].

The picture described here is, of course, a conjecture, which needs to be tested. For example, if a realistic case is found in which the $\mathcal{A}_*/4$ bound is violated by the contributions from both the interior and exterior light sheets, then we would have to modify the framework, e.g., by adding an extra $\mathcal{A}_*/4$ degrees of freedom on the holographic space. It is interesting that there is no known system that requires such a modification.

We finally discuss the connection with AdS/CFT. In the limit that the spacetime becomes asymptotically AdS, the location of the holographic screen is sent to spatial infinity, so that $\mathcal{A}_* \to \infty$. This implies that there are $N_* = e^{\mathcal{A}_*/4} \to \infty$ microstates for any spacetime configuration in D_{σ_*} for a leaf σ_*, including the case that it is a portion of the empty AdS space. Wouldn't this contradict the statement that the vacuum of a CFT is unique?

As we will discuss in Section 6.5, the degrees of freedom associated with N_* correspond to a freedom of sending information into the system at a later stage of time evolution, i.e. that of inserting operators at locations other than the point $x_{-\infty}$ corresponding to $\tau = -\infty$ on the conformally compactified AdS boundary. It is with this freedom that the CFT corresponds to the AdS limit of our theory including the N_* degrees of freedom:

$$\text{CFT} \iff \lim_{\mathcal{M} \to \text{asymptotic AdS}} \mathcal{T}, \tag{6.86}$$

where $\mathcal{M}$ is the spacetime inside the holographic screen, and $\mathcal{T}$ represents the theory under consideration. Here, we have taken the holographic screen to stay outside the cutoff surface (corresponding to the ultraviolet cutoff of the CFT) which is also sent to infinity.

This implies that if we want to consider a setup in which the evolution of the state is "autonomous" within the bulk, then we need to fix a configuration of operators at $x \neq x_{-\infty}$, i.e. we need to fully fix a boundary condition at the AdS boundary. The correspondence to our theory in this case is written as

$$\text{autonomous CFT} \iff \lim_{\mathcal{M} \to \text{asymptotic AdS}} \mathcal{T} \Big/ N_*. \tag{6.87}$$

[16] This bound was anticipated earlier [142] based on more phenomenological considerations.

The conventional vacuum state of the CFT corresponds to a special configuration of the N_* degrees of freedom that does not send in any signal to the system at later times (the simple reflective boundary conditions at the AdS boundary). Given the correspondence between the N_* degrees of freedom and boundary operators, we expect that this configuration is unique. The state corresponding to the CFT vacuum in our theory is then unique: the vacuum state of the theory $\mathcal{T}/N_*$ with the configuration of the N_* degrees of freedom chosen uniquely as discussed above.

Time evolution

Another feature of the holographic theory of general spacetimes beyond AdS/CFT is that the boundary space changes in time. This implies that we need to consider the theory in a large Hilbert space containing states living in different boundary spaces, Eq. (6.5). For states representing FRW universes, the relevant space can be written as

$$\mathcal{H} = \sum_{\mathcal{A}} \mathcal{H}_{\mathcal{A}}, \tag{6.88}$$

where $\mathcal{A}$ is the area of the leaf, and the sum of the Hilbert spaces is defined by Eq. (6.6).[17] While the microscopic theory involving time evolution is not yet available, we can derive its salient features by assuming that it reproduces the semiclassical time evolution in appropriate regimes. Here we discuss this issue for both direct sum and Russian doll structures. In particular, we consider a semiclassical time evolution in which a state having the leaf area $\mathcal{A}_1$ evolves into that having the leaf area $\mathcal{A}_2$ ($> \mathcal{A}_1$).

Direct sum structure

In this case there is a priori no need to introduce non-linearity in the algebra of observables, so we may assume that time evolution is described by a standard unitary operator acting on $\mathcal{H}$. In particular, time evolution of a state in $\mathcal{H}_{\mathcal{A}_1}$ into that in $\mathcal{H}_{\mathcal{A}_2}$ is given by a linear map from elements of $\mathcal{H}_{\mathcal{A}_1}$ to those in $\mathcal{H}_{\mathcal{A}_2}$.

Consider microstates $|\Psi_i^w\rangle$ ($i = 1, \cdots, e^{\mathcal{A}_1/4}$) representing the FRW universe with w when the leaf area is $\mathcal{A}_1$, $|\Psi_i^w\rangle \in \mathcal{H}_{\mathcal{A}_1,w} \subset \mathcal{H}_{\mathcal{A}_1}$; see Eq. (6.68). Assuming that all these states follow the standard semiclassical time evolution,[18] their evolution is given by

$$|\Psi_i^w\rangle \to |\Phi_i^w\rangle, \tag{6.89}$$

[17] More precisely, $\mathcal{H}_{\mathcal{A}}$ contains states whose leaf areas fall in the range between $\mathcal{A}$ and $\mathcal{A} + \delta\mathcal{A}$. The precise choice of $\delta\mathcal{A}$ is unimportant unless it is exponentially small in $\mathcal{A}$. For example, the dimension of $\mathcal{H}_A$ is $e^{\mathcal{A}/4}\delta\mathcal{A}$, so that the entropy associated with it is $\mathcal{A}/4 + \ln\delta\mathcal{A}$, which is $\mathcal{A}/4$ at the leading order in $1/\mathcal{A}$ expansion.

[18] Here we ignore the possibility that the equation of state changes between the two times, e.g., by a conversion of the matter content or vacuum decay. This does not affect our discussion below.

where $\{|\Phi_i^w\rangle\}$ is a subset of the microstates $|\Phi_j^w\rangle$ $(j = 1, \cdots, e^{\mathcal{A}_2/4})$ representing the FRW universe with w when the leaf area is $\mathcal{A}_2$, $|\Phi_j^w\rangle \in \mathcal{H}_{\mathcal{A}_2,w} \subset \mathcal{H}_{\mathcal{A}_2}$. This has an important implication. Suppose that the initial state of the universe is given by

$$|\Psi\rangle = \sum_i a_i |\Psi_i^w\rangle. \tag{6.90}$$

As we discussed before, if the effective number of terms in the sum is of order $e^{\mathcal{A}_1/4}$, namely if there are $e^{\mathcal{A}_1/4}$ nonzero a_i's with size $|a_i| \sim e^{-\mathcal{A}_1/8}$, then the state $|\Psi\rangle$ is not semiclassical, i.e. a firewall state (because a superposition of that many microstates changes the structure of the entanglement entropies). After the time evolution, however, this state becomes

$$|\Psi\rangle \to |\Phi\rangle = \sum_i a_i |\Phi_i^w\rangle, \tag{6.91}$$

where the number of terms in the sum is $e^{\mathcal{A}_1/4}$ because of the linearity of the map. This implies that the state $|\Phi\rangle$ is *not* a firewall state, since the number of terms is much (exponentially) smaller than the dimensionality of $\mathcal{H}_{\mathcal{A}_2,w}$: $e^{\mathcal{A}_1/4} \ll e^{\mathcal{A}_2/4}$. In particular, the state $|\Phi\rangle$ represents the standard semiclassical FRW universe with the equation of state parameter w.

This shows that this picture has a "built-in" mechanism of eliminating firewalls through time evolution, at least when the leaf area increases in time as we focus on here. This process happens very quickly—any macroscopic increase of the leaf area makes the state semiclassical regardless of the initial state.

Spacetime equals entanglement

In this case, time evolution from states in $\mathcal{H}_{\mathcal{A}_1}$ to those in $\mathcal{H}_{\mathcal{A}_2}$ is expected to be non-linear. Consider microstates $|\Psi_i^w\rangle$ $(i = 1, \cdots, e^{\mathcal{A}_1/4})$ representing the FRW universe with w when the leaf area is $\mathcal{A}_1$, $|\Psi_i^w\rangle \in \mathcal{H}_{\mathcal{A}_1}$. As before, requiring the standard semiclassical evolution for all the microstates, we obtain

$$|\Psi_i^w\rangle \to |\Phi_i^w\rangle, \tag{6.92}$$

where $\{|\Phi_i^w\rangle\}$ is a subset of the microstates $|\Phi_j^w\rangle$ $(j = 1, \cdots, e^{\mathcal{A}_2/4})$ representing the FRW universe with w when the leaf area is $\mathcal{A}_2$, $|\Phi_j^w\rangle \in \mathcal{H}_{\mathcal{A}_2}$. Suppose the initial state

$$|\Psi\rangle = \sum_i a_i |\Psi_i^w\rangle \equiv |\Psi^{w'}\rangle, \tag{6.93}$$

represents the FRW universe with $w' < w$. This is possible if the effective number of terms in the sum is of order $e^{\mathcal{A}_1/4}$, i.e. if there are $e^{\mathcal{A}_1/4}$ nonzero a_i's with size $|a_i| \sim e^{-\mathcal{A}_1/8}$. Now, if the time evolution map were linear, then this state would evolve into

$$|\Psi^{w'}\rangle \to |\Phi\rangle = \sum_i a_i |\Phi_i^w\rangle. \tag{6.94}$$

This state, however, is not a state representing the FRW universe with w', since the effective number of terms in the sum, $e^{\mathcal{A}_1/4}$, is exponentially smaller than $e^{\mathcal{A}_2/4}$, the required number to obtain a state with w' from the microstates $|\Phi_i^w\rangle$. To avoid this problem, the map from $\mathcal{H}_{\mathcal{A}_1}$ into $\mathcal{H}_{\mathcal{A}_2}$ must be non-linear so that $|\Psi^{w'}\rangle$ evolves into $|\Phi^{w'}\rangle$ containing $e^{\mathcal{A}_2/4}$ terms when expanded in terms of $|\Phi_i^w\rangle$.

Here we make two comments. First, the non-linearity of the map described above does not necessarily mean that the time evolution of semiclassical degrees of freedom (given as excitations on the background states considered here) is non-linear, since the definition of these degrees of freedom would also be non-linear at the fundamental level. In fact, from observation this evolution must be linear, at least with high accuracy. This requirement gives a strong constraint on the structure of the theory. Second, the non-linearity seen above arises when the area of the boundary space changes, $\mathcal{A}_1 \to \mathcal{A}_2 \neq \mathcal{A}_1$. Since the area of the boundary is fixed in the AdS/CFT limit (with the standard regularization and renormalization procedure), this non-linearity does not show up in the CFT time evolution, generated by the dilatation operator with respect to the $t = -\infty$ point in the compactified Euclidean space.[19]

We finally discuss relations between different $\mathcal{H}_B$'s. While we do not know how they are related, for example they could simply exist as a direct sum in the full Hilbert space $\mathcal{H} = \bigoplus_B \mathcal{H}_B$, an interesting possibility is that their structure is analogous to the Russian doll structure within a single $\mathcal{H}_B$. Specifically, let us introduce the notation to represent the Russian doll structure as

$$\{|\Psi^w\rangle\} \prec \{|\Psi^{w'}\rangle\} \quad \text{for} \quad w' < w, \tag{6.95}$$

meaning that the left-hand side is a measure zero subset of the closure of the right-hand side. We may imagine that states $|\Psi_B\rangle$ representing spacetimes with boundary B and states $|\Psi_{B'}\rangle$ representing those with boundary B' are related similarly as

$$\{|\Psi_B\rangle\} \prec \{|\Psi_{B'}\rangle\} \quad \text{for} \quad \|B\| < \|B'\|. \tag{6.96}$$

(The relation may be more complicated; for example, some of the $|\Psi_B\rangle$'s are related with $|\Psi_{B'}\rangle$'s and some with $|\Psi_{B''}\rangle$'s with $B'' \neq B'$.) Ultimately, all states in realistic (cosmological) spacetimes may be related with those in asymptotically Minkowski space as

$$\{|\Psi_B\rangle\} \prec \{|\Psi_{B'}\rangle\} \cdots \prec \{|\Psi_{\text{Minkowski}}\rangle\}, \tag{6.97}$$

since the boundary area for asymptotically Minkowski space is infinity, $\mathcal{A}_{\text{Minkowski}} = \infty$. Does string theory formulated in an asymptotically Minkowski background (using S-matrix elements) correspond to the present theory as

$$\text{String theory} \iff \lim_{\mathcal{M}\to\text{asymptotic Minkowski}} \mathcal{T}\,? \tag{6.98}$$

[19] This does not mean that the interior of a black hole is described by state-independent operators in the CFT. It is possible that the CFT does not provide a description of the black hole interior; see discussion in Section 6.4.

Here, the $\mathcal{T}/N_{\text{Minkowski}}$ portion is described by the scattering dynamics, and the $N_{\text{Minkowski}}$ degrees of freedom are responsible for the initial conditions, where $N_{\text{Minkowski}} = e^{\mathcal{A}_{\text{Minkowski}}/4}$; see the next section. If this is indeed the case, then it would be difficult to obtain a useful description of cosmological spacetimes directly in that formulation, since they would correspond to a special measure zero subset of the possible asymptotic states.

6.5 Discussion

In this final section, we discuss some of the issues that have not been addressed in the construction so far. This includes the possibility of sending signals from a past singularity or past null infinity (in the course of time evolution) and the interpretation of a closed universe in which the area of the leaf changes from increasing to decreasing once the scale factor at the leaf starts decreasing. We argue that these issues are related to that of "selecting a state"—even if the theory is specified we still need to provide selection conditions on a state, usually given in the form of boundary conditions (e.g. initial conditions). Our discussion here is schematic, but it allows us to develop intuition about how quantum gravity in general spacetimes might work at the fundamental level.

Signals from a past singularity or past null infinity

As mentioned in Section 6.4, the evolution of a state in the present framework is not fully autonomous. Consistent with the covariant entropy bound, we may view a holographic state to carry the information on the two (future-directed ingoing and past-directed outgoing) light sheets associated with the leaf it represents. However, this is not enough to determine a future state because there may be signals sent into the system from a past singularity or past null infinity (signals originating from the lower right direction between the two 45° lines in Fig. 6.13).

To be specific, let us consider a (not necessarily FRW) universe beginning with a big bang. As shown in Fig. 6.14, obtaining a future state (represented by the upper 45° line) in general requires a specification of signals from the big bang singularity, in addition to the information contained in the original state (the lower 45° line). We usually avoid this issue by requiring the "cosmological principle," i.e. spatial homogeneity and isotropy, which determines what conditions one must put at the singularity—with this requirement, the state of the universe is determined by the energy density content in the universe at a time. Imposing this principle, however, corresponds to choosing a very special state. This is because there is no reason to expect that signals sent from the singularity at different times τ (defined holographically) are correlated in such a way that the system appears as homogeneous and isotropic in some time slicing. In fact, this was one of the original motivations for inflationary cosmology [75, 115, 6].

In some cases, appropriate conditions can be obtained by assuming that the spacetime under consideration is a portion of larger spacetime. For example, if the universe is dominated

by negative curvature at an early stage, it may arise from bubble nucleation [44], in which case the homogeneity and isotropy would result from the dynamics of the bubble nucleation [76]. Even in this case, however, we would still need to specify similar conditions in the ambient space in which our bubble forms, and so on. More generally, the analysis here says that to obtain a future state, we need to specify the information coming from the directions tangent to the past-directed light rays. This, however, is morally the same as the usual situation in physics in which we need to specify boundary (e.g. initial) conditions beyond the dynamical laws the system obeys.

The situation is essentially the same in the limit of AdS/CFT; we only need to consider the AdS boundary instead of the big bang singularity. To obtain future states, it is not enough to specify the initial state, given by a local operator inserted at the point $x_{-\infty}$ corresponding to $\tau = -\infty$ on the conformally compactified AdS boundary. We also have to specify other (possible) boundary operators inserted at points other than $x_{-\infty}$.

String theory formulated in terms of the S-matrix deals with this issue by adopting an asymptotically Minkowski time slice in which all the necessary information is viewed as being in the initial state. This, however, does not change the amount of information needed to specify the state, which is infinite in asymptotically Minkowski space (because one can in principle send an infinite amount of information into the system from past null infinity).

Closed universes—time in quantum gravity

Consider a closed universe in which the vacuum energy is negligible throughout its history. In such a universe, the area of the leaf changes from increasing to decreasing in the middle of its evolution. On the other hand, we expect that the area of the leaf for a "generic" state increases monotonically, since the number of independent states representing spacetime with the leaf area $\mathcal{A}$ goes as $e^{\mathcal{A}/4}$. What does this imply?

We interpret that states representing universes like these are "fine-tuned," so that they do not obey the usual second law of thermodynamics as applied to the Hilbert space of quantum gravity. This does not mean that they are meaningless states to consider. Rather, it means that we need to scrutinize carefully the concept of time in quantum gravity.

There are at least three different views of time in quantum gravity; see, e.g., Ref. [131]. First, since time parameterization in quantum gravity is nothing other than a gauge choice, the state $|\tilde{\Psi}\rangle$ of the *full* system—whatever its interpretation—satisfies the constraint [48, 188]

$$H|\tilde{\Psi}\rangle = 0, \tag{6.99}$$

where H is the time evolution operator, in our context the generator of a shift in τ. In this sense, the concept of time evolution does not apply to the full state $|\tilde{\Psi}\rangle$.[20] However, this of course does not mean that *physical time* we perceive is nonexistent. Time we observe can

[20]Reference [48] states that Eq. (6.99) need not apply in an infinite world; for example, the state of the system $|\Psi_\infty\rangle$ may depend on time in asymptotically Minkowski space. We view that Eq. (6.99) still applies in this case by interpreting $|\tilde{\Psi}\rangle$ to represent the full system, including the "exterior" degrees of freedom discussed in Section 6.4 (the degrees of freedom corresponding to $N_{\text{Minkowski}}$ below Eq. (6.98)) as well as

be defined as correlations between subsystems (e.g. between an object playing the role of a clock and the rest) [48, 149], at least in some branch of $|\tilde{\Psi}\rangle$. Another way to define time is through probability flow in $|\tilde{\Psi}\rangle$. Suppose $|\tilde{\Psi}\rangle$ is expanded in a set of states $|\Psi_i\rangle$ each of which represents a well-defined semiclassical spacetime when such an interpretation is applicable:

$$|\tilde{\Psi}\rangle = \sum_i c_i |\Psi_i\rangle. \tag{6.100}$$

According to the discussion in Section 6.4, $|\Psi_i\rangle$'s are approximately orthogonal in the appropriate limit, and the constraint in Eq. (6.99) implies

$$\sum_j c_j U_{ij} = c_i, \qquad U_{ij} \equiv \langle \Psi_i | e^{-iH\delta\tau} | \Psi_j \rangle, \tag{6.101}$$

where U_{ij} is (effectively) unitary

$$\sum_j U_{ij} U^*_{kj} = \sum_j U_{ji} U^*_{jk} = \delta_{ik}. \tag{6.102}$$

Multiplying Eq. (6.101) with its conjugate and using Eq. (6.102), we obtain

$$\begin{aligned} 0 = &- |c_i|^2 \sum_{j\neq i} |U_{ji}|^2 + \sum_{j\neq i} |c_j|^2 |U_{ij}|^2 \\ &+ \sum_{j\neq i} c_i c_j^* U_{ii} U^*_{ij} + \sum_{j\neq i} c_j c_i^* U_{ij} U^*_{ii} + \sum_{\substack{j,k\neq i \\ j\neq k}} c_j c_k^* U_{ij} U^*_{ik}. \end{aligned} \tag{6.103}$$

In the regime where the WKB approximation is applicable, the terms in the second line are negligible compared with those in the first line because of a rapid oscillation of the phases of $c_{j,k}$'s, so that

$$|c_i|^2 \sum_{j\neq i} |U_{ji}|^2 = \sum_{j\neq i} |c_j|^2 |U_{ij}|^2, \tag{6.104}$$

implying that the "current of probability" is conserved. We may regard this current as time flow. The time defined in this way—which we call *current time*—need not be the same as the physical time defined through correlations, although in many cases the former agrees approximately with the latter or the negative of it (up to a trivial shift and rescaling).

In a closed universe (with a negligible vacuum energy), it is customary to impose the boundary condition

$$c_i = 0 \quad \text{for} \quad \{|\Psi_i\rangle \,|\, a = 0\}, \tag{6.105}$$

i.e. the wavefunction vanishes when the scale factor goes to zero [48]. With this boundary condition, current time τ flows in a closed circuit. The direction of the flow agrees with that

the "interior" degrees of freedom represented by $|\Psi_\infty\rangle$. The time evolution of $|\Psi_\infty\rangle$ is then understood as correlations between the interior and exterior degrees of freedom, as described below.

of physical time in the branches where $da/d\tau > 0$, while the two are exactly the opposite in the branches where $da/d\tau < 0$. (The latter statement follows, e.g., from the analysis in Ref. [4], which shows that given a lower entropy final condition the most likely history of a system is the CPT conjugate of the standard time evolution.) Our time evolution in earlier sections concerns the flow of current time. The (apparent) violation of the second law of thermodynamics then arises because the condition of Eq. (6.105) selects a special, "standing wave" solution from the viewpoint of the current time flow. This is, however, not a fine-tuning from the point of view of the quantum theory in a similar way as the electron energy levels of the hydrogen atom are not regarded as fine-tuned states.

The fact that current time flows toward lower entropy states does not mean that a physical observer living in the $da/d\tau < 0$ phase sees a violation of the second law of thermodynamics. Since the whole system evolves as time reversal of a standard entropy increasing process, *including memory states of the observer*, a physical observer always finds the evolution of the system to be the standard one [133, 4]; in particular, he/she always finds that the universe is expanding.

Static quantum multiverse—selecting the state in the landscape

The analysis of string theory suggests that the theory has a multitude of metastable vacua each of which leads to a distinct low energy effective theory [34, 104, 168, 53]. Combining this with the fact that many of these vacua lead to inflation (which is future eternal at the semiclassical level) leads to the picture of the inflationary multiverse [78, 185, 117, 116]. The picture suggests that our universe is one of many bubble universes, and it cannot be a closed universe that will eventually collapse as the one discussed above. How is the state of the multiverse selected then?

A naive semiclassical picture implies that the state of the multiverse evolves asymptotically into a superposition of supersymmetric Minkowski worlds and (possibly) "singularity worlds" resulting from the big crunches of AdS bubble universes [133]. This is because any other universe is expected to decay eventually. There are basically two possibilities for the situation in a full quantum theory.

The first possibility is that the multiverse is in a "scattering state." This essentially preserves the semiclassical intuition. From the viewpoint of the current time flow, the multiverse begins as an asymptotic state, experiences nontrivial cosmology at an intermediate stage, and then dissipates again into the asymptotic Minkowski and singularity worlds. In the earlier stage of the evolution in which the coarse-grained entropy decreases in τ, the directions of current and physical time flows are opposite, while in the later stage of increasing entropy, the flows of the two times are in the same direction. The resulting picture is similar to that of Ref. [41]: the multiverse evolves asymptotically into both forward and backward in (current) time. This, however, does not mean that a physical observer, who is a part of the system, sees an entropy decreasing universe; the observer always finds that his/her world obeys the second law of thermodynamics.

A problem with this possibility is that specifying the theory of quantum gravity, e.g. the structure of the Hilbert space and Hamiltonian, is not enough to obtain the state of the multiverse and hence make predictions. We would need a separate theory to specify initial conditions. Furthermore, having a lower course-grained entropy at the turn-around point (the point at which the coarse-grained entropy changes from decreasing to increasing in the current time evolution) requires a more carefully chosen initial condition. This leads to the issue of understanding why we are "ordinary observers," carrying course-grained entropies (much) smaller than that needed to have any consciousness—a variant of the well-known Boltzmann brain problem [56, 5, 147] (the argument applied to space of initial conditions, rather than to a thermal system).

The alternative, and perhaps more attractive, possibility is that the multiverse is in a "bound state" [134]. Specifically, the multiverse is in a *normalizable* state satisfying the constraint of Eq. (6.99) (as well as any other constraints):

$$|\tilde{\Psi}\rangle = \sum_i c_i |\Psi_i\rangle; \qquad \sum_i |c_i|^2 < \infty. \tag{6.106}$$

This is a normalization condition in spacetime, rather than in space as in usual quantum mechanics, and it allows us to determine, in principle, the state of the multiverse once the theory is given.[21] As in the case of a collapsing closed universe, current time flows in a closed circuit(s) to the extent that this concept is applicable. This suggests that the multiverse does not probe an asymptotic supersymmetric Minkowski region or the big crunch singularity of an AdS bubble. The origin of this phenomenon must be intrinsically quantum mechanical as it contradicts the naive semiclassical picture. In fact, such a situation is not new in physics. As is well known, the hydrogen atom cannot be correctly described using classical mechanics: any orbit of the electron is unstable with respect to the emission of synchrotron radiation. The situation in the quantum multiverse may be similar—quantum mechanics is responsible for the very existence of the system.

Once the state of the multiverse is determined, we should be able to use it to give predictions or explanations. This requires us to develop a prescription for extracting answers to physical questions about the state. The prescription would certainly involve coarse-graining (as one cannot even store the information of all possible microstates of the multiverse within the multiverse), and it should reproduce the standard Born rule giving probabilistic predictions in the appropriate regime. Perhaps, the normalization condition of Eq. (6.106) is required in order for this prescription to be well-defined.

[21] If there are multiple solutions $|\tilde{\Psi}_I\rangle$, it is natural to assume that the multiverse is in the maximally mixed state $\rho = \frac{1}{N}\sum_{I=1}^{N} |\tilde{\Psi}_I\rangle\langle\tilde{\Psi}_I|$ (in the absence of more information). Here, we have taken $|\tilde{\Psi}_I\rangle$'s to be orthonormal.

6.6 Appendix for Chapter 6

Spacelike Monotonicity Theorem

Let H be a past holographic screen, foliated by compact marginally anti-trapped surfaces i.e. leaves, $\{\sigma_r\}$. Here, r is a (non-unique) real parameter taken to be a monotonically increasing function of the leaf area. For each leaf we can construct the two future-directed null vector fields (up to overall normalization) and denote them k^a and l^a, which satisfy

$$\theta_k = 0, \qquad \theta_l > 0. \tag{6.107}$$

Now let h^a a leaf-orthogonal vector field tangent to H and normalized by the condition $h^a \partial_a r = 1$. Note that h^a must point in the direction of increasing area. We can always put $h^a = \alpha l^a + \beta k^a$ for some smooth real-valued functions α and β on H. The Bousso-Engelhardt area theorem implies that $\alpha > 0$ everywhere. There is no restriction on the sign of β: it can even have indefinite sign on a single leaf.

Let A_r be a $d-2$ dimensional region in a leaf σ_r and let ∂A_r denote its boundary, where d is the spacetime dimension. This region can be transported to a region $A_{r'}$ in a nearby leaf $\sigma_{r'}$ by following the integral curves of the leaf-orthogonal vector field h^a. While Ref. [159] pointed out that $\|A_r\|$ is an increasing function of r, this by itself does not guarantee that $S(A_r)$ monotonically increases. Nonetheless, we now show that $S(A_r)$ indeed monotonically increases if h^a is spacelike.

Theorem 4. *Suppose that H is a past holographic screen foliated by leaves $\{\sigma_r\}$ and assume that the parameter r is oriented to increase as leaf area increases. Assume that H is spacelike on some particular leaf which we take to be σ_0 by shifting r if necessary. Let A_0 be a subregion of σ_0 and define $A_r \subset \sigma_r$ by transporting points in A_0 along the integral curves of the leaf-orthogonal vector field in H. Then, $S(A_r)$ is a monotonically increasing function of r.*

Proof. Let h^a be the leaf-orthogonal vector field tangent to H with $h^a \partial_a r = 1$ and note that $h^a\big|_{\sigma_0}$ is spacelike. The compactness of σ_0 now allows us to find $r_0 > 0$ such that $h^a\big|_{H[-r_0,r_0]}$ is spacelike. Here we have introduced the convenient notation

$$H[r_1, r_2] = \bigcup_{r_1 \le r \le r_2} \sigma_r. \tag{6.108}$$

In what follows, we will assume that the extremal surface $E(A_r)$ anchored to ∂A_r deforms smoothly as a function of r at $r = 0$. If this is not the case, a phase transition occurs at $r = 0$ which will give rise to a discontinuity in the derivative of $S(A_r)$. However, we can then note that our theorem applies at r slightly greater than zero (where H is still spacelike and where no phase transition occurs) and also at r slightly smaller than zero. This implies that $S(A_r)$ is monotonically increasing at $r = 0$ even if $E(A_r)$ "jumps" at $r = 0$ so that the derivative of $\|E(A_r)\|$ has a discontinuity.

The maximin construction of $E(A_0)$ ensures that there exists $\Sigma_0 \in \mathcal{C}_{\sigma_0}$ such that $E(A_0) = \min(A_0, \Sigma_0)$. Here, $\mathcal{C}_\sigma$ denotes the collection of all complete codimension-1 achronal surfaces lying in D_σ that are anchored to σ, and $\min(A, \Sigma)$ denote the $d-2$ dimensional surface of minimal area lying in Σ that is homologous to A. If $0 < \epsilon < r_0$, let

$$\Sigma_\epsilon = \Sigma_0 \cup H[0, \epsilon]. \tag{6.109}$$

We claim that $\Sigma_\epsilon \in \mathcal{C}_{\sigma_\epsilon}$ for small ϵ. First we check that Σ_ϵ is achronal. Since Σ_0 and $H[0, \epsilon]$ are achronal independently, we focus on their intersection at σ_0. The definition of $\mathcal{C}_{\sigma_0}$ requires that Σ_0 lies in D_{σ_0} so that a vector pointing from σ_0 to Σ_0 has the form $c_1 k^a - c_2 l^a$ with $c_1, c_2 > 0$. Meanwhile, a vector pointing from σ_0 to $H[0, \epsilon]$ is proportional to $h^a\big|_{\sigma_0} = |\alpha| l^a - |\beta| k^a$. Here we have made use of the fact that $\alpha > 0$ and $\beta < 0$ for a spacelike past holographic screen. We see now that Σ_0 lies "inside" σ_0 while h^a points toward the "outside." This ensures that Σ_ϵ is achronal for sufficiently small ϵ. All that is left to check is that Σ_ϵ lies inside of D_{σ_ϵ}. But this is clear because a vector pointing from σ_ϵ toward Σ_ϵ is proportional to $-h^a\big|_{\sigma_\epsilon} = -|\alpha| l^a + |\beta| k^a$ which is indeed directed into D_{σ_ϵ}. That $\Sigma_\epsilon \in \mathcal{C}_{\sigma_\epsilon}$ is now clear for small ϵ.

We now construct an ϵ-dependent family of $d-2$ dimensional surfaces lying on Σ_0 that are anchored to ∂A_0, which we will denote by Ξ_ϵ. Begin by fixing a small ϵ with $0 < \epsilon < r_0$ and defining a projection function $\pi_\epsilon : H[0, \epsilon] \to \sigma_0$ in the natural way: if $p \in H[0, \epsilon]$, follow the integral curves of h^a, starting from p, until a point in σ_0 is reached. The result is $\pi_\epsilon(p)$. We can now define Ξ_ϵ:

$$\Xi_\epsilon = \Big(\min(A_\epsilon, \Sigma_\epsilon) \cap \Sigma_0 \Big) \bigcup \pi_\epsilon \Big(\min(A_\epsilon, \Sigma_\epsilon) \cap H[0, \epsilon] \Big). \tag{6.110}$$

If ϵ is sufficiently small, the fact that $H[0, \epsilon]$ has a positive definite metric, along with the fact that $E(A_0)$ is not tangent to σ_0 anywhere, ensures that $\|\pi_\epsilon\big(\min(A_\epsilon, \Sigma_\epsilon) \cap H[0, \epsilon]\big)\| < \|\min(A_\epsilon, \Sigma_\epsilon) \cap H[0, \epsilon]\|$. From this it follows that

$$\|\Xi_\epsilon\| < \|\min(A_\epsilon, \Sigma_\epsilon)\|. \tag{6.111}$$

On the other hand, because $\pi_\epsilon(\partial A_\epsilon) = \partial A_0$, we know that Ξ_ϵ is a codimension-2 surface anchored to ∂A_0 that lies only on Σ_0. Thus,

$$4S(A_0) = \|\min(A_0, \Sigma_0)\| \leq \|\Xi_\epsilon\|. \tag{6.112}$$

Noting that the maximin construction of $E(A_\epsilon)$ requires

$$\|\min(A_\epsilon, \Sigma_\epsilon)\| \leq 4S(A_\epsilon), \tag{6.113}$$

we find $S(A_0) < S(A_\epsilon)$. □

Qubit Model

Model and applications to quantum gravity

Here we describe a toy model for holographic states representing FRW universes, presented originally in Ref. [143]. We consider a Hilbert space for N ($\gg 1$) qubits $\mathcal{H} = (\mathbf{C}^2)^{\otimes N}$. Let Δ ($\leq N$) be a nonnegative integer and consider a typical superposition of 2^Δ product states

$$|\Psi\rangle = \sum_{i=1}^{2^\Delta} a_i \, |x_1^i x_2^i \cdots x_N^i\rangle, \tag{6.114}$$

where $\{a_i\}$ is a normalized complex vector, and $x^i_{1,\cdots,N} \in \{0,1\}$. Given an integer n with $1 \leq n < N$, we can break the Hilbert space into a subsystem Γ for the first n qubits and its complement $\bar{\Gamma}$. We are interested in computing the entanglement entropy S_Γ of Γ.

Suppose $n \leq N/2$. If $\Delta \geq n$, then i in Eq. (6.114) runs over an index that takes many more values than the dimension of the Hilbert space for Γ, so that Page's argument [144] tells us that Γ has maximal entanglement entropy: $S_\Gamma = n \ln 2$. On the other hand, if $\Delta < n$ then the number of terms in Eq. (6.114) is much less than both the dimension of the Hilbert space of Γ and that of $\bar{\Gamma}$, which limits the entanglement entropy: $S_\Gamma = \Delta \ln 2$. We therefore obtain

$$S_\Gamma = \begin{cases} n & n \leq \Delta, \\ \Delta & n > \Delta, \end{cases} \tag{6.115}$$

for $\Delta < N/2$, while

$$S_\Gamma = n, \tag{6.116}$$

for $\Delta \geq N/2$. Here and below, we drop the irrelevant factor of $\ln 2$. The value of S_Γ for $n > N/2$ is obtained from $S_\Gamma = S_{\bar{\Gamma}}$ since $|\Psi\rangle$ is pure.

The behavior of S_Γ in Eqs. (6.115, 6.116) models that of $S(\gamma)$ in Section 6.3. The correspondence is given by

$$\frac{n}{N} \leftrightarrow \frac{\|\Gamma\|}{\mathcal{A}_*}, \tag{6.117}$$

$$\frac{\Delta}{N} \leftrightarrow \frac{1}{2} Q_w\left(\frac{\pi}{2}\right), \tag{6.118}$$

for $\Delta \leq N/2$.[22] The identification of Eq. (6.117) is natural if we regard the $N = \mathcal{A}_*/4$ qubits as distributing over a leaf σ_* with each qubit occupying a volume of 4 in Planck units. The quantity Δ controls what universe a state represents. For fixed Δ, different choices of the product states $|x_1^i x_2^i \cdots x_N^i\rangle$ and the coefficients a_i give e^N independent microstates for the FRW universe with $w = f(\Delta/N)$. The function f is determined by Eq. (6.118); in particular, $f = -1$ (> -1) for $\Delta/N = 1/2$ ($< 1/2$).

[22] States with $\Delta > N/2$ cannot be discriminated from those with $\Delta = N/2$ using S_Γ alone. Below, we only consider the states with $N/4 \leq \Delta \leq N/2$.

This model can be used to argue for features of the holographic theories discussed in Section 6.4. We consider two cases:

Direct sum structure — In this case, each of the subspaces $\mathcal{H}_{*,w}$ is modeled by the N qubit system described here. Consider $\mathcal{H}_{*,w}$ with a fixed w. States representing the FRW universe with w then encompass e^N independent microstates in this space. These microstates form "effective vector space" in that a superposition involving less than $e^{O(\delta w N)}$ of them leads only to another microstate representing the same FRW universe with w. (We say that these states comprise "fat" preferred axes.) Most of the states in $\mathcal{H}_{*,w}$, containing more than $e^{O(\delta w N)}$ of the w microstates, are regarded as non-semiclassical, i.e. firewall or unphysical, states.

Russian doll structure — In this case, the entire $\mathcal{H}_*$ space is modeled by the N qubits, and the states representing various FRW universes are all elements of this single Hilbert space of dimension e^N. An important point is that the set of states with *any* fixed Δ_w provide a complete basis for the whole Hilbert space, where $\Delta_w \equiv N f^{-1}(w)$. This implies that we can obtain a state with any $w' < w$ by superposing $e^{\Delta_{w'} - \Delta_w}$ states with Δ_w, and we can also obtain a state with $w' > w$ as a superposition of carefully chosen e^{Δ_w} states with Δ_w. We call this the "Russian doll" structure, which is depicted schematically in Fig. 6.11.

Effective incoherence of superpositions

We now focus on the latter case and consider a normalized superposition

$$|\Psi\rangle = c_1|\Psi_1\rangle + c_2|\Psi_2\rangle, \tag{6.119}$$

of two states

$$|\Psi_1\rangle = \sum_{i=1}^{2^{\Delta_1}} a_i \, |x_1^i x_2^i \cdots x_N^i\rangle \qquad \left(\sum_{i=1}^{2^{\Delta_1}} |a_i|^2 = 1\right), \tag{6.120}$$

$$|\Psi_2\rangle = \sum_{i=1}^{2^{\Delta_2}} b_i \, |y_1^i y_2^i \cdots y_N^i\rangle \qquad \left(\sum_{i=1}^{2^{\Delta_2}} |b_i|^2 = 1\right), \tag{6.121}$$

with $\Delta_1 \neq \Delta_2$ and

$$\Delta_1, \Delta_2 \leq \frac{N}{2}. \tag{6.122}$$

Here, the coefficients a_i and b_i are random, as are the binary values $x^i_{1,\cdots,N}$ and $y^i_{1,\cdots,N}$, and $|c_1|^2 + |c_2|^2 = 1$ up to an exponentially suppressed correction arising from $\langle\Psi_1|\Psi_2\rangle \neq 0 \lesssim O(2^{-|\Delta_1 - \Delta_2|/2})$. We are interested in the reduced density matrix

$$\rho_{1\cdots n} = \mathrm{Tr}_{n+1\cdots N}\, \rho, \tag{6.123}$$

obtained by performing a partial trace on

$$\rho = |\Psi\rangle\langle\Psi| = |c_1|^2|\Psi_1\rangle\langle\Psi_1| + |c_2|^2|\Psi_2\rangle\langle\Psi_2| + c_1c_2^*|\Psi_1\rangle\langle\Psi_2| + c_2c_1^*|\Psi_2\rangle\langle\Psi_1|, \tag{6.124}$$

over the subsystem consisting of the first n qubits. We will only consider the case where $n < N/2$.

We begin our analysis by considering $\mathrm{Tr}_{n+1\cdots N}|\Psi_1\rangle\langle\Psi_1|$. It is convenient to write

$$|\Psi_1\rangle\langle\Psi_1| = \sum_{i=1}^{2^{\Delta_1}} |a_i|^2\, |x_1^i\cdots x_N^i\rangle\langle x_1^i\cdots x_N^i| + \sum_{\substack{i,j=1\\ i\neq j}}^{2^{\Delta_1}} a_i a_j^*\, |x_1^i\cdots x_N^i\rangle\langle x_1^j\cdots x_N^j|. \tag{6.125}$$

Upon performing the partial trace over $|\Psi_1\rangle\langle\Psi_1|$, the first sum gives a diagonal contribution to the reduced density matrix

$$D_{11} = \sum_{i=1}^{2^{\Delta_1}} |a_i|^2\, |x_1^i\cdots x_n^i\rangle\langle x_1^i\cdots x_n^i|. \tag{6.126}$$

The second sum gives a correction

$$\tilde{D}_{11} = \sum_{\substack{i,j=1\\ i\neq j}}^{2^{\Delta_1}} a_i a_j^*\, |x_1^i\cdots x_n^i\rangle\langle x_1^j\cdots x_n^j|\, \delta_{x_{n+1}^i, x_{n+1}^j}\cdots\delta_{x_N^i, x_N^j}. \tag{6.127}$$

We now consider two cases:

(i) $\Delta_1 > n$.
Because $2^{\Delta_1} \gg 2^n$, it is clear that D_{11} is a $2^n \times 2^n$ diagonal matrix with every diagonal entry approximately given by

$$\frac{2^{\Delta_1}}{2^n}\left\langle |a_i|^2\right\rangle = 2^{-n}. \tag{6.128}$$

(Note that $\langle |a_i|^2\rangle = 2^{-\Delta_1}$ because $|\Psi_1\rangle$ is normalized and random.) Thus, D_{11} is a fully mixed state. Now observe that $\tilde{D}_{11}$ consists of almost all zeros. In fact, looking at Eq. (6.127) we see that there are $2^{2\Delta_1-N+n}$ nonzero entries of average absolute value $2^{-\Delta_1}$. Given that $\Delta_1 \leq N/2$, we conclude that $\tilde{D}_{11}$ has exponentially fewer nonzero entries than D_{11}, and that each nonzero entry has exponentially smaller size than the entries of D_{11}.

(ii) $\Delta_1 \leq n$.
In this case, D_{11} is a diagonal matrix having 2^{Δ_1} nonzero entries of order $2^{-\Delta_1}$. The number of nonzero entries in $\tilde{D}_{11}$ is, again, $2^{2\Delta_1-N+n}$, each having the average absolute value $2^{-\Delta_1}$. The effect of $\tilde{D}_{11}$ is highly suppressed because its number of nonzero entries

is exponentially smaller than that of D_{11}. In fact, for the number of nonzero entries in $\tilde{D}_{11}$ to compete with that in D_{11}, we would need $2\Delta_1 - N + n \geq \Delta_1$, which, however, mean

$$\Delta_1 \geq N - n > \frac{N}{2}, \tag{6.129}$$

a contradiction.

Summarizing, $\mathrm{Tr}_{n+1\cdots N}|\Psi_1\rangle\langle\Psi_1| = D_{11} + \tilde{D}_{11}$ is a diagonal matrix having $2^{\min\{\Delta_1,n\}}$ nonzero entries of order $2^{-\min\{\Delta_1,n\}}$, up to exponentially suppressed effects. The same analysis obviously applies to $\mathrm{Tr}_{n+1\cdots N}|\Psi_2\rangle\langle\Psi_2| = D_{22} + \tilde{D}_{22}$ with $\Delta_1 \to \Delta_2$.

We now turn our attention to the matrix $\mathrm{Tr}_{n+1\cdots N}|\Psi_1\rangle\langle\Psi_2|$, which we denote as $\tilde{D}_{12}$:

$$\tilde{D}_{12} = \sum_{i=1}^{2^{\Delta_1}} \sum_{j=1}^{2^{\Delta_2}} a_i b_j^* \, |x_1^i \cdots x_n^i\rangle\langle y_1^j \cdots y_n^j| \delta_{x_{n+1}^i, y_{n+1}^j} \cdots \delta_{x_N^i, y_N^j}. \tag{6.130}$$

We argue, along similar lines to the above, that $\tilde{D}_{12}$ is exponentially smaller than $|c_1|^2 D_{11} + |c_2|^2 D_{22}$, unless $|c_1|$ or $|c_2|$ is exponentially suppressed. Once again, we have several cases:

(i) $\Delta_1, \Delta_2 \leq n$.
In this case, $|c_1|^2 D_{11} + |c_2|^2 D_{22}$ is a diagonal matrix having 2^{Δ_1} nonzero entries of order $2^{-\Delta_1}$ and 2^{Δ_2} nonzero entries of order $2^{-\Delta_2}$. Considering Eq. (6.130), $\tilde{D}_{12}$ consists of zeros except for $2^{\Delta_1+\Delta_2-N+n}$ nonzero entries with the average absolute value $\left\langle |a_i b_j^*| \right\rangle = 2^{-(\Delta_1+\Delta_2)/2}$. The number of these entries, however, is exponentially smaller than 2^{Δ_1}, since having $\Delta_1 + \Delta_2 - N + n \geq \Delta_1$ would require $\Delta_2 \geq N - n > N/2$; similarly, it is also exponentially smaller than 2^{Δ_2}. Moreover the changes of the exponentially rare eigenvalues affected are at most of $O(1)$. We conclude that the effect of $\tilde{D}_{12}$ is exponentially suppressed.

(ii) $\Delta_1, \Delta_2 > n$.
In this case, the condition that $|c_1|^2 + |c_2|^2 = 1$ ensures that $|c_1|^2 D_{11} + |c_2|^2 D_{22}$ is a $2^n \times 2^n$ unit matrix multiplied by 2^{-n}. Meanwhile, $\tilde{D}_{12}$ consists of zeros except for $2^{\Delta_1+\Delta_2-N+n} \ll 2^n$ nonzero entries of size $2^{-(\Delta_1+\Delta_2)/2} \ll 2^{-n}$.

(iii) $\Delta_1 \leq n < \Delta_2$.
In this case, D_{22} is a $2^n \times 2^n$ unit matrix multiplied by 2^{-n} while D_{11} is a diagonal matrix having 2^{Δ_1} nonzero entries of order $2^{-\Delta_1}$. Once again, the number of nonzero entries in $\tilde{D}_{12}$ is exponentially smaller than 2^{Δ_1}, since $\Delta_1 + \Delta_2 - N + n \geq \Delta_1$ would require $\Delta_2 \geq N - n > N/2$, and the fractional corrections to eigenvalues from these entries are of order $2^{-(\Delta_2-n)}$. This implies that the effect of $\tilde{D}_{12}$ is negligible. The same argument also applies to the case that $\Delta_2 \leq n < \Delta_1$.

We conclude that for $n < N/2$, we find

$$\rho_{1\cdots n} = |c_1|^2 D_{11} + |c_2|^2 D_{22} = \sum_{i=1}^{2^{\Delta_1}} |a_i|^2 \, |x_1^i \cdots x_n^i\rangle\langle x_1^i \cdots x_n^i| + \sum_{i=1}^{2^{\Delta_2}} |b_i|^2 \, |y_1^i \cdots y_n^i\rangle\langle y_1^i \cdots y_n^i|, \tag{6.131}$$

up to effects exponentially suppressed in $N \approx O(\mathcal{A}_*)$. This implies that the reduced density matrix for the state $|\Psi\rangle$ takes the form of an incoherent classical mixture

$$\rho_{1\cdots n} = |c_1|^2 \rho^{(1)}_{1\cdots n} + |c_2|^2 \rho^{(2)}_{1\cdots n}, \tag{6.132}$$

where $\rho^{(k)}_{1\cdots n} = \mathrm{Tr}_{n+1\cdots N}|\Psi_k\rangle\langle\Psi_k|$ $(k = 1, 2)$ are the reduced density matrices we would obtain if the state were $|\Psi_k\rangle$.

The form of Eq. (6.131) also implies that the entanglement entropy

$$S_{1\cdots n} = -\mathrm{Tr}_{1\cdots n}(\rho_{1\cdots n} \ln \rho_{1\cdots n}), \tag{6.133}$$

obeys a similar linear relation

$$S_{1\cdots n} = |c_1|^2 S^{(1)}_{1\cdots n} + |c_2|^2 S^{(2)}_{1\cdots n} + O(1), \tag{6.134}$$

unless $|c_1|$ or $|c_2|$ is exponentially small. Here, $S^{(k)}_{1\cdots n} = -\mathrm{Tr}_{1\cdots n}(\rho^{(k)}_{1\cdots n} \ln \rho^{(k)}_{1\cdots n})$. This can be seen by considering the same three cases as above. If $\Delta_1, \Delta_2 \leq n$, $\rho_{1\cdots n}$ is a diagonal matrix having 2^{Δ_1} nonzero entries with average value $|c_1|^2 2^{-\Delta_1}$ and 2^{Δ_2} nonzero entries with average value $|c_2|^2 2^{-\Delta_2}$. In this case,

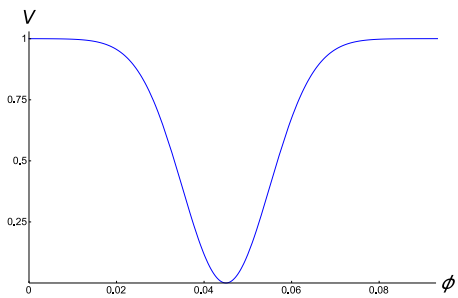
$$S_{1\cdots n} = -|c_1|^2 \ln \frac{|c_1|^2}{2^{\Delta_1}} - |c_2|^2 \ln \frac{|c_2|^2}{2^{\Delta_2}} = |c_1|^2 \Delta_1 \ln 2 + |c_2|^2 \Delta_2 \ln 2 + O(1), \tag{6.135}$$

while we have $S^{(k)}_{1\cdots n} = \Delta_k \ln 2$. The $O(1)$ correction from linearity is the entropy of mixing, given by

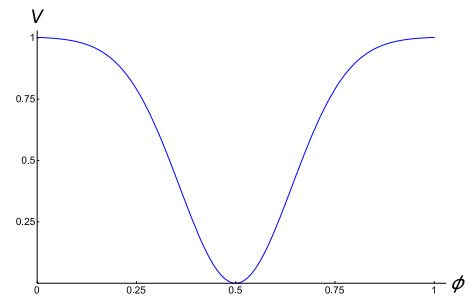
$$S_{1\cdots n,\mathrm{mix}} = -|c_1|^2 \ln |c_1|^2 - |c_2|^2 \ln |c_2|^2. \tag{6.136}$$

If $\Delta_1, \Delta_2 > n$, then $\rho_{1\cdots n}$ is a unit matrix multiplied by 2^{-n}. From this it follows that $S_{1\cdots n} = n \ln 2 = |c_1|^2 n \ln 2 + |c_2|^2 n \ln 2$, which is desirable given that $S^{(k)}_{1\cdots n} = n \ln 2$ for $\Delta_k > n$.[23] Finally, if $\Delta_1 < n < \Delta_2$, $\rho^{(1)}_{1\cdots n}$ has 2^{Δ_1} nonzero entries of mean value $2^{-\Delta_1}$ while $\rho^{(2)}_{1\cdots n}$ is a unit matrix multiplied by 2^{-n}. Because $2^{-\Delta_1} \gg 2^{-n}$ the total density matrix $\rho_{1\cdots n}$ given by Eq. (6.131) is diagonal and has 2^{Δ_1} entries of size $|c_1|^2 2^{-\Delta_1}$ and 2^n entries of size $|c_2|^2 2^{-n}$. We thus find that $S_{1\cdots n} = |c_1|^2 \Delta_1 \ln 2 + |c_2|^2 n \ln 2 + S_{1\cdots n,\mathrm{mix}} = |c_1|^2 S^{(1)}_{1\cdots n} + |c_2|^2 S^{(2)}_{1\cdots n} + O(1)$. (This expression is valid for $\Delta_1 = n < \Delta_2$ as well.)

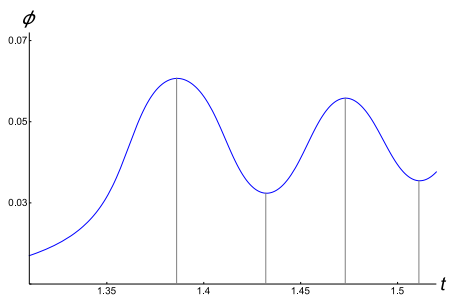
[23] The absence of the mixing contribution in this case is an artifact of the specific qubit model considered here, arising from the fact that two universes cannot be discriminated unless n is larger than one of $\Delta_{1,2}$; see Eq. (6.115). In realistic cases, the mixing contribution should always exist for any macroscopic region in the holographic space as two different universes can be discriminated in that region; see, e.g., Fig. 6.5.



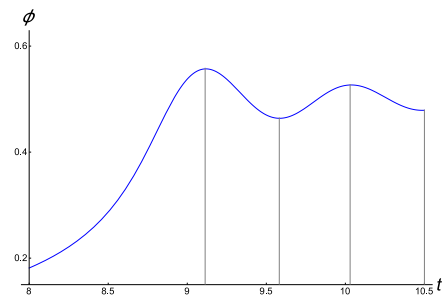
(a) Steep Potential



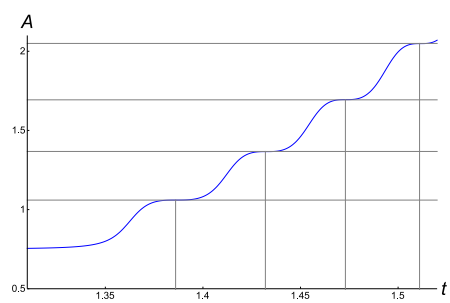
(b) Broad Potential



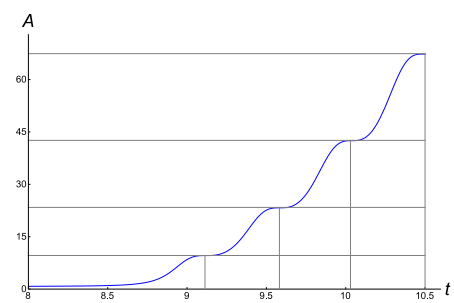
(c) $\phi(t)$ for the steep potential.



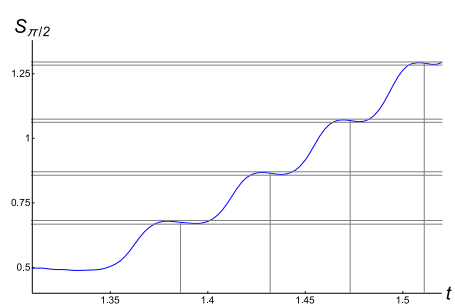
(d) $\phi(t)$ for the broad potential.



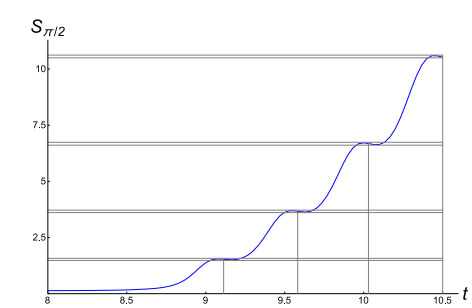
(e) $\|L(\pi/2, t)\|$ for the steep potential.



(f) $\|L(\pi/2, t)\|$ for the broad potential.



(g) $S(\pi/2, t)$ for the steep potential.



(h) $S(\pi/2, t)$ for the broad potential.

Figure 6.10: A steep potential (a) leading to the time evolution of the scalar field (b), the area of a leaf hemisphere (c), and the screen entanglement entropy (d). The same for a broad potential (e)–(h).

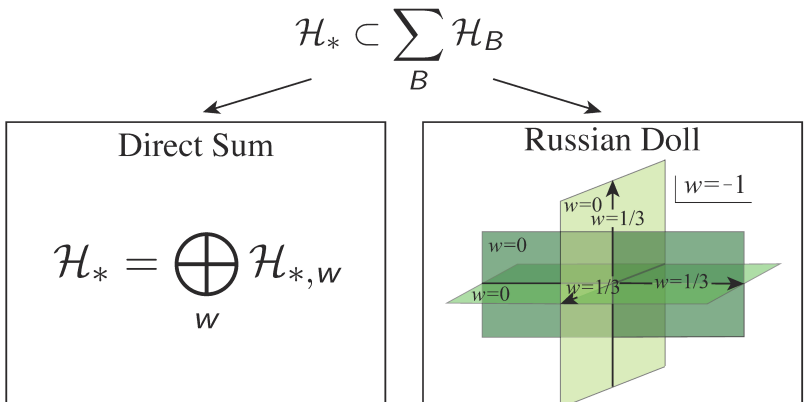


Figure 6.11: Possible structures of the Hilbert space $\mathcal{H}_*$ for a fixed boundary space B. In the direct sum structure (left), each semiclassical spacetime in D_{σ_*} has its own Hilbert space $\mathcal{H}_{*,w}$. The Russian doll structure (right) corresponds to the scenario of "spacetime equals entanglement," i.e. the entanglement entropies of the holographic degrees of freedom determine spacetime in D_{σ_*}. This implies that a superposition of exponentially many semiclassical spacetimes can lead to a different semiclassical spacetime.

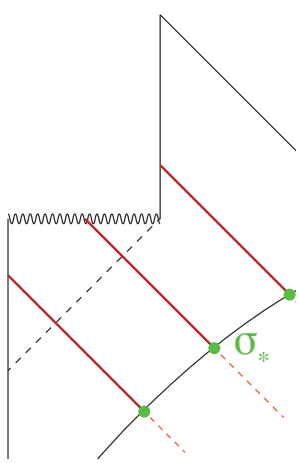


Figure 6.12: If a black hole forms inside the holographic screen, future-directed ingoing light rays emanating orthogonally from the leaf σ_* at an intermediate time may hit the singularity before reaching a caustic. While the diagram here assumes spherical symmetry for simplicity, the phenomenon can occur more generally.

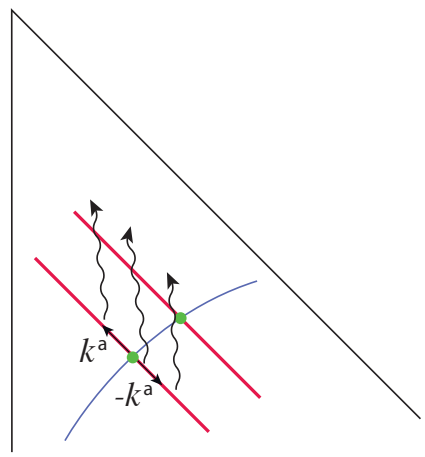


Figure 6.13: To determine a state in the future, we need information on the "exterior" light sheet, the light sheet generated by light rays emanating from σ_* in the $-k^a$ directions, in addition to that on the "interior" light sheet, i.e. the one generated by light rays emanating in the $+k^a$ directions.

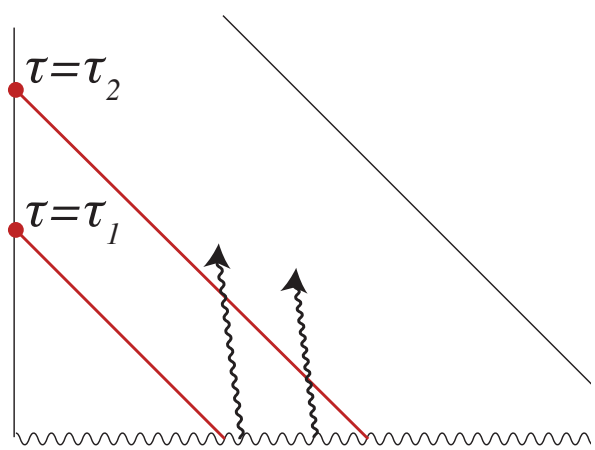


Figure 6.14: In a universe beginning with a big bang, obtaining a future state requires a specification of signals from the big bang singularity, in addition to the information contained in the original state. In an FRW universe this is done by imposing spatial homogeneity and isotropy, which corresponds to selecting a fine-tuned state from the viewpoint of the big bang universe.

Chapter 7

The Boundary Dual of Bulk Local Operators

7.1 Introduction

The emergence of bulk locality in the AdS/CFT correspondence [122, 189] has yet to receive a satisfactory explanation in terms of the behavior of holographic CFTs. While gravity prohibits exact locality in a quantum theory, when the gravitational coupling is sufficiently small, local physics must be a good approximation in the bulk. There should be a manifestation of this "emergence of locality" in the boundary theory.

One way to tackle this issue is by studying the ways in which bulk degrees of freedom are encoded in the CFT. It is thus natural to ask if there is a boundary dual of local bulk fields in the regime where semiclassical field theory holds. While the extrapolate dictionary [18] states that bulk fields at spacelike infinity are dual to local operators on the boundary, points deep in the bulk require a nonlocal holographic description. There are many well-known ways to reconstruct bulk fields in terms of nonlocal boundary operators [18, 80, 103, 94] with support in a variety of boundary regions. All of these procedures, however, require solving bulk equations of motion which presupposes knowledge of the bulk spacetime. If one were not explicitly told the metric in the bulk, is there any way to determine whether or not a given operator is bulk local? To put this question differently, is the concept of a local bulk operator in any way distinguished in the boundary theory?

The primary goal of this work is to address this question. We will find that a powerful tool to this end is the concept of subregion duality. The notion that a boundary domain of dependence should be thought of as being dual to some region of the bulk, which originally arose from considerations of causal wedge reconstruction, was made precise recently by [7, 102, 49] where it was concluded that a bulk operator can be reconstructed in a subregion of the CFT if and only if its support is contained in the entanglement wedge of that CFT region [49]. This conclusion was made in the context of a new development in AdS/CFT: the role of quantum error correction. It is now understood that a semiclassical bulk spacetime

description is associated with a code subspace of the boundary Hilbert space, and that various inequivalent boundary reconstructions of bulk operators become equivalent when restricted to the code subspace.

This modern form of subregion duality will reveal a novel characterization of locality in the bulk. Given a holographic CFT and a code subspace dual to some unknown geometry, we will provide a procedure that can identify, up to certain caveats, whether or not an operator is dual to a bulk local operator. As a byproduct of our method, we are also able to reconstruct the causal structure (equivalently, the metric up to a conformal factor) of a large region in the bulk. In some examples, this region can penetrate event horizons.

Outline. We start, in section 7.2, by reviewing the arguments and motivation for the quantum error correcting view in holography. In particular, we sketch the proof of [49] that a bulk operator is reconstructable in a boundary region if and only if its support is contained entirely in the entanglement wedge of that boundary region.

Section 7.3 contains the major constructions of this work. We define the notion of a *superficially local operator* without making direct reference to the bulk. These are bulk operators that are "as local as the boundary can directly tell." Their defining characteristic is the great variety of boundary regions in which they can be reconstructed. In a certain region of the bulk called the *localizable region*, operators are local if and only if they are superficially local. However, there are situations in which superficially local operators correspond to nonlocal bulk operators that are supported outside of the localizable region. The bulk regions in which these problematic operators lie will be referred to as *clumps*. Fortunately, clumps appear to always be identifiable from the boundary theory because they are associated with phase transitions. Thus, they can be identified and thrown away, leaving only the superficially local operators that are authentically dual to bulk local operators.

The set of superficially local operators can be given an equivalence relation by identifying two operators when they can be reconstructed in exactly the same boundary regions. After removing clumps, the set of equivalence classes of superficially local operators is naturally identified with the bulk localizable region.

In 7.4, we note that the commutation relations amongst these operators reveals the causal structure in the localizable region. Thus, we are able to reconstruct the metric in this portion of the bulk up to a conformal rescaling. This approach is similar at heart to that of [61] where a bulk reconstruction is accomplished by means of light-cone cuts. We argue, in fact, that there are numerous interesting connections between our approach and that involving cut singularities.

7.2 Principles of Subregion Duality

This section provides a brief review of the quantum error correcting view of AdS/CFT. Readers already familiar with the conclusions of [7, 49] may wish to proceed to section 7.3

There is a zoo of different methods for expressing bulk fields in terms of CFT operators.

The extrapolate dictionary [18] gives a precise relationship between limiting values of bulk fields and CFT operators with corresponding scaling dimensions. It is also possible to express operators lying deeper in the bulk in therms of CFT quantities by solving equations of motion in the bulk [80, 18, 103, 128, 94]. Of these approaches, one of relevance for our considerations is the causal wedge reconstruction, which generalizes the Rindler reconstruction of [80]. This prescription expresses local bulk fields in terms of CFT operators localized to a special boundary subregion. Specifically, if R is region in the boundary with domain of dependence $D^\partial(R)$, and if $\mathrm{CW}(R) = J^+[D^\partial(R)] \cap J^-[D^\partial(R)]$ is the causal wedge [98] of R, then causal wedge reconstruction allows a bulk field in $\mathrm{CW}(R)$ to be expressed as a smeared operator in $D^\partial(R)$.[1]

Causal wedge reconstruction suggests the possibility that subregions in the boundary are enough to understand the physics of associated bulk subregions. However, despite what is suggested from the analysis of [80], the causal wedge is, in general, not the largest possible region that a boundary subregion holographically describes in the semiclassical limit. Instead, the bulk region dual to a CFT region R is the entanglement wedge of R, denoted by $\mathrm{EW}(R)$ [187, 93, 101]. $\mathrm{EW}(R)$ can be defined as follows. Let Σ be a spacelike bulk surface that, after conformal compactification of M, is a Cauchy slice for the unphysical bulk spacetime. Require that Σ contains R and its HRT surface $\mathrm{ext}\, R$. Let S denote the part of Σ between R and $\mathrm{ext}\, R$. The domain of dependence of S (computed in the unphysical spacetime) is the entanglement wedge of R. It is known that $\mathrm{EW}(R) \supseteq \mathrm{CW}(R)$ [93]. As we review below, [49] gave a precise sense in which a boundary region R should really be thought of as being dual to its entanglement wedge. This is the most refined and powerful known form of "subregion duality" [36] in AdS/CFT.

Before discussing entanglement wedge reconstruction, we note that subregion duality, even in the form of [80], raises major puzzles [7]. For example, an operator $\phi(p)$ deep within the bulk can be taken to lie in many different causal wedges. Thus, a causal wedge reconstruction of the form

$$\phi(p) = \int_{D^\partial(R)} K(p,x)O(x)dx \tag{7.1}$$

manifestly commutes with all operators in the complement region $\bar{R}$. This argument can be repeated for many different boundary regions and used to show that a bulk field $\phi(p)$ near the center of AdS can be written in a way that manifestly commutes with any given operator in the boundary. This directly implies what should have been obvious: that each choice of reconstruction for $\phi(p)$ is a different operator in the CFT. This is not an inconsistency. Various reconstructions of $\phi(p)$ are distinct CFT operators, but the CFT Hilbert space is much larger[2] than the Hilbert space relevant for a bulk operator on a spacetime background. The explanation of the multitude of distinct CFT operators is therefore that there is a special

[1] The smearing function has to be understood in a distributional sense. For details see [128, 36]. Such subtleties will not be important for what follows.

[2] The basic concept that semiclassical excitations give rise to exponentially small subspaces of a Hilbert space describing quantum gravitational physics has played a role in many related areas. See, e.g., [142, 136, 151]

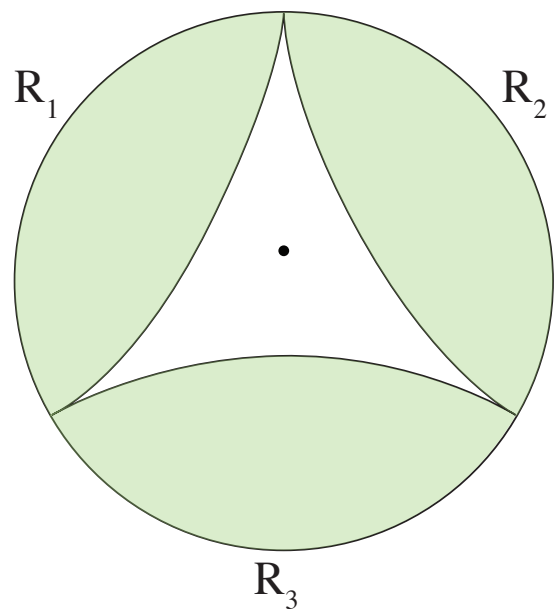


Figure 7.1: The operator depicted in the center of this figure is not in CW(R_1), CW(R_2), or CW(R_3). However, it does lie in the causal wedge of the union of any two regions CW($R_i \cup R_j$) and can thus be written in terms of boundary operators in the algebra of the combined regions.

subspace of the Hilbert space, the code subspace, which describes the states that $\phi(p)$ is defined on. The restriction of all reconstructions of $\phi(p)$ to this subspace reproduce $\phi(p)$. This is a quantum-error correcting property of the CFT: the action of different operators defined in different regions is the same when restricting to special subspaces called *code subspaces*.

The necessity for such a redundant descriptions of bulk operators was made particularly obvious with the following argument [7] illustrated in figure 7.1 . Consider a partition the boundary into 3 equal regions $R_1, R_2,$ and R_3 which only have points on their boundaries in common. Taking the vacuum state for simplicity, their causal wedges will not contain points that are close to the center of the bulk spacetime. Thus, there is no HKLL smearing over any one region that reconstructs a local bulk operator near the center. However, the causal wedge of the union of any two regions CW($R_i \cup R_j$) does contain the bulk point of interest and the HKLL procedure can be used. The different choices cannot represent the same CFT operator, since their support is on causally disconnected regions.[3]

Review of the DHW argument

[3] The mutual intersection actually includes points on the boundaries of the R_i. However, repeating the argument with slightly different regions circumvents the possibility that the reconstruction of ϕ is achieved only in the algebra of ∂R_i

The fact that the entanglement wedge EW(R) is the "largest" bulk region that can be reconstructed from the algebra of R will play a critical role in our work. For this reason, we will briefly review the arguments in [7, 49], focusing especially on the aspects of this literature that will be the most relevant for the framework that we begin to develop in section 7.3.

Suppose that we are given[4] a particular code subspace $G \subset H$ which is known to be a span of states obtained by acting with a small number of low energy operators on a state where a semiclassical bulk exists; in particular, within G, gravitational backreaction of bulk fields can be treated perturbatively. Dong, Harlow, and Wall (DHW) proved that if the support of an operator ϕ is contained in EW(R), then that operator can be reconstructed in R [49]. This means that there is an element of the algebra of R whose action on states in the code subspace is the same as the action of ϕ.

To understand the proof given in [49], we first refer to a result from quantum information. Refs. [7, 25, 24] show that if we have a code subspace G and some factorization of the full Hilbert space $G \subset H_R \otimes H_{\bar{R}}$, and if ϕ is some operator that acts within G (it's action send states in the code subspace to other states in the code subspace), then the following two statements are equivalent.

1. There exists an operator O_R on H_R such that for any $|\psi\rangle \in G$,
$$\phi|\psi\rangle = O_R|\psi\rangle \qquad \phi^\dagger|\psi\rangle = O_R^\dagger|\psi\rangle. \tag{7.2}$$
2. For any operator $X_{\bar{R}}$ on on $H_{\bar{R}}$, we have
$$[\phi, X_{\bar{R}}]\big|_G = 0. \tag{7.3}$$

While this theorem follows purely from quantum information, it plays a critical role in the entanglement wedge reconstruction argument. As suggested by the notation, we will associate R with the factorization induced from boundary regions and G will be a code subspace with a semiclassical bulk interpretation. We can now discuss [49], which establishes that bulk semiclassical operators satisfy condition 7.3, and the reconstructability follows because this is equivalent to 7.2.

We know the boundary Hilbert space can be factorized into a region and its complement $H = H_R \otimes H_{\bar{R}}$. For states with a semiclassical bulk interpretation, we can think about the extremal surface anchored to ∂R as inducing its own tensor factorization of the code subspace $\mathrm{G}_{\mathrm{EW}(R)} \otimes G_{\mathrm{EW}(\bar{R})}$.

Consider two states $|\psi_0\rangle, |\psi_1\rangle \in G$ and the reduced density matrices obtained by tracing out the appropriate complement regions in the two factorizations
$$\begin{aligned} \rho^0_{\bar{R}} &= \mathrm{Tr}_R|\psi_0\rangle\langle\psi_0| \\ \rho^0_{\mathrm{EW}(\bar{R})} &= \mathrm{Tr}_{\mathrm{EW}(R)}|\psi_0\rangle\langle\psi_0| \end{aligned} \tag{7.4}$$

[4]While we take the code subspace as given, it should be possible to identify code subspaces purely from the CFT. For example, a necessary (but not sufficient) condition for a collection of states to lie in the same code subspace is that the collection has the property that subregions have entanglement entropies differing only by sub-leading contributions in N.

Similarly, the density matrices $\rho^1_{\bar{R}}$ and $\rho^1_{\mathrm{EW}(\bar{R})}$ are defined by the state $|\psi_1\rangle$.

The statement of a theorem in [49] is that if the states satisfy:

$$\rho^0_{\mathrm{EW}(\bar{R})} = \rho^1_{\mathrm{EW}(\bar{R})} \implies \rho^0_{\bar{R}} = \rho^1_{\bar{R}} \tag{7.5}$$

then, an operator of the form $\phi = \mathbb{1} \otimes \phi_{\mathrm{EW}(R)}$ acting only within the entanglement wedge of R will satisfy the two equivalent properties of 7.2 and 7.3.

To understand this, we note that the result in [102] established a precise relationship between the bulk and boundary modular hamiltonian. This provides the connection between the first equality and second equality in 7.5. Now, the operator ϕ supported in the entanglement wedge of a boundary region R does not affect the state in the complement wedge (this just follows from semiclassical field theory). Thus, if we define $|\psi_1\rangle$ as

$$|\psi_1\rangle = e^{i\epsilon\phi}|\psi_0\rangle \tag{7.6}$$

the first equality in 7.5 is satisfied. The second equality then implies that the expectation value of any operator in the algebra of $\bar{R}$ is the same in both states:

$$\langle\psi_0|X_{\bar{R}}|\psi_0\rangle - \langle\psi_1|X_{\bar{R}}|\psi_1\rangle = 0 \tag{7.7}$$

Rewriting the second term using 7.6 and expanding to first order in ϵ we obtain 7.3.

This proves that within the code subspace, we can express operators in the entanglement wedge of R in terms of operators in the algebra of R. Moreover, if an operator on G has support outside $\mathrm{EW}(R)$, it must have no reconstruction in R. To see this, suppose that an operator ϕ on G had support outside $\mathrm{EW}(R)$ so that it fails to commute with some operator ϕ' on $\mathrm{EW}(\bar{R})$. The argument above shows that there exists a reconstruction $O'_{\bar{R}}$ of ϕ' that acts on $\bar{R}$. If ϕ could be reconstructed with an operator O_R on R, we would have $[O'_{\bar{R}}, O_R] = 0$ which contradicts the fact that $[\phi', \phi] \neq 0$.

Our final conclusion is that an operator acting on a code subspace can be reconstructed in a region R of the CFT if and only if its support in entirely contained in $\mathrm{EW}(R)$. By exploiting the reconstructability for states in the code subspace, we now explore how the bulk, including the conformal metric, is encoded in the CFT.

We note that the reconstructability argument itself is a statement about a special class of quantum states and makes no reference to the plank length in the bulk. However, in making the connection between the reduced density matrix in the entanglement wedge [102] and the boundary, one clearly needs to assume some notion of locality. In particular, this involves taking $N \to \infty$.

7.3 Superficially Local Operators

For the rest of this paper we work in the context of the "infinite N limit." It is assumed that there are code subspaces $\{G\}$ of the CFT Hilbert space H that are holographically dual

to quantum field theory on (asymptotically AdS) spacetime backgrounds. Setting $N = \infty$ in this way may cause discomfort, especially with some of the more complicated things we discuss below, and for this reason we have provided appendix .1 which defines our quantities while taking the large N limit more carefully. Even without reading the appendix, the majority of our development can made much more precise simply by replacing equalities with approximate equalities which, in the large N limit, approach authentic equalities.

In this section we are going to almost completely answer a fundamental question: Suppose that a code subspace G is given and that we are told that G is dual to some unknown field theory on some unknown spacetime background. Let ϕ be a given operator on G. Is ϕ dual to a local operator? Note that we are given no information about ϕ (other than how it acts on G) and, in particular, it is probably not a local CFT operator. The ability to answer this question is equivalent to finding all of the CFT operators that are dual to local bulk operators with respect to our particular code subspace.

Prior work addresses related issues but falls short of providing a general identification of local bulk operators. Consider, again, the HKLL method [80]. If ϕ is a quantum field in the bulk M, then, given a point $p \in M$, it is possible to solve the field equation of motion and obtain an expression of the form

$$\phi(p) = \left(\int_{\partial M} K(p,x) O(x) d^{D-1}x \right) \Big|_G . \tag{7.8}$$

Here, the boundary field O is the one associated with ϕ through the extrapolate dictionary. As discussed above, the integration kernel K is not unique. While different choices of K yield different CFT operators, the restriction of these different choices of operators to the code subspace G must always give the same answer.

At a first glance, equation 7.8 appears to not only identify the nonlocal CFT operators that are dual to local bulk operators, but even provides a formula for them. This is not the case however. The integration kernel can only be found by solving equations of motion on the curved spacetime background M, and this assumes knowledge of what the background is. There are very few code subspaces for which the corresponding geometry is known. Another reason that the HKLL procedure is unsatisfactory for our purposes is that it only identifies a subset of the boundary operators that are dual to local bulk operators. We would like to find a more general characterization of locality in the bulk at leading order in $1/N$.

Comparing Locality of Operators

Our guiding principle is that that, roughly speaking, the more local a bulk operator ϕ is, the more distinct boundary regions exist for which ϕ can be reconstructed. This follows from subregion duality as explained in section 7.2. To make this concept more precise, we are going to employ the full power of the quantum error-correcting structure of AdS/CFT to introduce a function $\mathcal{Q}$ that maps operators on G to the collection of all possible boundary regions that can reconstruct a given operator. $\mathcal{Q}$ will then provide a measure of locality of every operator. We now explain this precisely.

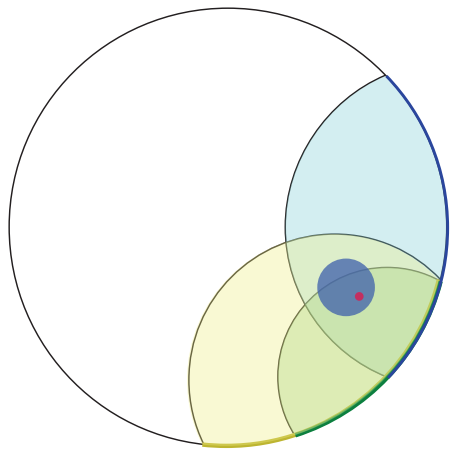


Figure 7.2: A nonlocal bulk operator ϕ_1 will clearly lie in fewer regions than an operator ϕ_2 whose support is entirely contained in the first $\mathcal{Q}(\phi_1) \subset \mathcal{Q}(\phi_2)$.

Let $\mathcal{R}$ denote the collection of all $D-2$ dimensional achronal submanifolds of ∂M. Informally, $\mathcal{R}$ is the collection of all regions R upon which one would compute a von Neumann entropy by anchoring stationary surfaces [157, 99] to ∂R. Note that we are not restricting to a single time slice of ∂M. If $R \in \mathcal{R}$ and ϕ is an operator that, along with its hermitian conjugate, acts on the code subspace G, then ϕ is said to be reconstructable in R if there exists O in the algebra of R such that $O\big|_G = \phi$ and $O^\dagger\big|_G = \phi^\dagger$. We now give a critical definition:

Definition 1. *Suppose that ϕ is an operator on G and $R \in \mathcal{R}$. Then, we define*

$$\mathcal{Q}(\phi) = \{R \in \mathcal{R} \,\big|\, \phi \textit{ is reconstructable in } R\}.$$

Whatever the (unknown) geometry of M is, subregion duality (see section 7.2) gives a geometrical condition for $\mathcal{Q}(\phi)$ to contain a region R. Specifically, $R \in \mathcal{Q}(\phi)$ if and only if the (bulk) support[5] of ϕ is contained in the entanglement wedge of R. This immediately implies the following properties of $\mathcal{Q}$:

Proposition 7.3.1. *Let ϕ_1 and ϕ_2 be two operators on the code subspace G. Then,*

1. *If* $\operatorname{supp}\phi_1 \supseteq \operatorname{supp}\phi_2$ *, then* $\mathcal{Q}(\phi_1) \subseteq \mathcal{Q}(\phi_2)$*,*
2. *if* $\operatorname{supp}\phi_1 = \operatorname{supp}\phi_2$*, then* $\mathcal{Q}(\phi_1) = \mathcal{Q}(\phi_2)$*.*

[5]The support of an operator is defined as follows. Let A be a (possibly nonlocal) operator on a quantum field theory on the curved spacetime M. Let U be the set of points in M such that for every point p in U, every local bulk operator at p commutes with A. Then, the support of A, denoted by supp A, is given by $M \setminus (J_+(U) \cup J_-(U))$

Note that the converses to these statements, though seemingly desirable, are false in many cases. This is somewhat disappointing: the bulk support of an operator is a property of the operator's bulk description while $\mathcal{Q}$ is a function that is manifestly defined in the boundary theory. Our goal is to find a "boundary-only" characterization of bulk locality, so we would be much better off if the converse to Proposition 7.3.1 were in fact true.

What $\mathcal{Q}$ does accomplish is that it identifies the support of an operator to the greatest possible "resolution" that the boundary theory can easily see. For this reason we define an equivalence relation on operators on G: $\phi_1 \sim \phi_2$ if $\mathcal{Q}(\phi_1) = \mathcal{Q}(\phi_2)$. We use the notation $[\phi]$ to denote the equivalence class of ϕ with respect to this relation. In other words, $[\phi] = Q^{-1}(Q(\phi))$. Two operators are in the same class if they are "the same as far as $\mathcal{Q}$ can tell." We can attempt to compare the locality of two operators by putting a partial ordering on the collection of equivalence classes by writing $[\phi_1] \leq [\phi_2]$ if $\mathcal{Q}(\phi_1) \subseteq \mathcal{Q}(\phi_2)$ (which is a well-defined relation). Note that a trivial operator like the identity on G, denoted by 1_G, can be reconstructed in any region. Thus, $[\phi] \leq [1_G]$ for any operator ϕ on G.

We are now ready to give a plausible characterization of a local bulk operator by means of $\mathcal{Q}$.

Definition 2. *Suppose that ϕ is an operator on G. ϕ is said to be* superficially local *if*

1. $[\phi] \neq [1_G]$ *and*
2. *Every operator ϕ' with the property that $[\phi] \leq [\phi']$ has $[\phi'] \in \{[\phi], [1_G]\}$.*

We emphasize that the definition of a superficially local operator makes reference only to the boundary theory. Thus, we can use this definition to offer an answer to the question posed above: if we are given a large N CFT with a Hilbert space H, a subspace G of H, and an operator ϕ, and if we told that G is a code subspace corresponding to an unknown bulk spacetime, then we can guess that ϕ is a local operator in the dual bulk theory if it acts on G and if its restriction to G is a superficially local operator. This answer turns out to be right in many cases.

The word "superficial" is used for two reasons. First, as we will shortly see, there are examples of asymptotically AdS spacetimes for which some local bulk operators (for instance, those lying close to a spacelike singularity) are not superficially local. Second, we will not prove that every superficially local operator is local in the bulk. The first of these deficiencies is completely unavoidable and it is tempting to contemplate its relation to the difficulties of using AdS/CFT to describe points deep within a black hole interior [10] (although we will not pursue such contemplations here). The second apparent deficiency is not a problem: in section 7.4 we will argue that it is possible to identify when a given equivalence class of superficially local operators contains operators that are not actually local in the bulk. This argument will be made in the boundary theory. The concept of superficial locality therefore provides a way to confidently identify a very large collection of operators on G that should be interpreted as local operators in the bulk. We now explain exactly which bulk operators can be found in this way.

The Localizable Region

As above, let M be the asymptotically AdS bulk spacetime that is dual to a code subspace G of a CFT in the large N limit with Hilbert space H. In this section we are going to identify a special subset of M, denoted by $\mathrm{Loc}(M)$, which has the property that local bulk operators at points in $\mathrm{Loc}(M)$ can be successfully identified in the boundary theory through the consideration of superficially local operators.

Definition 3. *The* localizable region *of M, denoted* $\mathrm{Loc}(M)$*, is the set of points $p \in M$ satisfying*

1. *If* $\mathrm{supp}\,\phi = \{p\}$*, then ϕ is superficially local and*
2. *if* $\mathrm{supp}\,\phi = \{p\}$ *and* $[\phi'] = [\phi]$*, then* $\mathrm{supp}\,\phi' = \{p\}$.

Elements of $\mathrm{Loc}(M)$ will sometimes be called *localizable points*. Note that $\mathrm{Loc}(M)$ is a subset of the bulk and its definition makes reference to the concept of the bulk support of an operator, so this definition is not particularly transparent from the boundary theory. However, a connection with the boundary theory becomes apparent when $\mathrm{Loc}(M) = M$:

Proposition 7.3.2. *If* $\mathrm{Loc}(M) = M$*, an operator ϕ on G is superficially local if and only if it is local in the bulk. Moreover, if ϕ_1 and ϕ_2 are two superficially local operators with* $[\phi_1] = [\phi_2]$*, then they must be local at the same bulk point.*

Proof. If ϕ is a local operator, the definition of $\mathrm{Loc}(M)$ immediately demands that ϕ is superficially local. Conversely, let suppose that ϕ is superficially local. If ϕ is not local in the bulk, then there are at least two distinct points p and q in the support of ϕ. Let ϕ' be a local operator at p. By Proposition 7.3.1, the fact that $\mathrm{supp}\,\phi' \subseteq \mathrm{supp}\,\phi$ means that $[\phi] \leq [\phi']$. But ϕ is superficially local and ϕ' is nontrivial so we conclude that $[\phi] = [\phi']$. The definition of the localizable region now demands that $\mathrm{supp}\,\phi = \{p\}$, a contradiction.

Now suppose that ϕ_1 and ϕ_2 are two superficially local operators with $[\phi_1] = [\phi_2]$. From what we just proved, we know that ϕ_1 is local at some point, so the definition of the localizable region immediately demands that ϕ_1 and ϕ_2 are local at the same point. □

This result is a first step to providing a boundary description of $\mathrm{Loc}(M)$ because the notion of superficial locality is one of the boundary theory. Unfortunately the hypothesis of Proposition 7.3.2 is often too much to ask for. To better understand this, consider the following result which which establishes a geometrical bulk interpretation of $\mathrm{Loc}(M)$.[6]

Theorem 5. $p \in \mathrm{Loc}(M)$ *if and only if there exists a subset $\mathcal{R}_0$ of the collection of boundary regions $\mathcal{R}$ such that*

$$\bigcap_{R \in \mathcal{R}_0} \mathrm{EW}(R) = \{p\}.$$

[6]Theorem 5 elucidates the connection between our program and the ideas of [47, 46, 17, 15, 45, 91]. Note this work is primarily interested in the reconstruction of bulk geometry while our focus is on operator reconstruction. However, below in section 7.4 we will reconstruct aspects of the bulk geometry.

Proof. Suppose first that there exists $\mathcal{R}_0$ satisfying the condition given in the statement of the theorem. Fix a local bulk operator ϕ at p so that $\operatorname{supp}\phi = \{p\}$. $\mathcal{Q}(\phi)$ must contain all regions R with $p \in \mathrm{EW}(R)$ so, in particular, $\mathcal{R}_0 \subseteq \mathcal{Q}(\phi)$. If ϕ' is some operator on G with $[\phi'] \geq [\phi]$, then $\mathcal{Q}(\phi) \subseteq \mathcal{Q}(\phi')$ so we have

$$\begin{aligned}\operatorname{supp}\phi' \subseteq \bigcap_{R\in\mathcal{Q}(\phi')} \mathrm{EW}(R) &\subseteq \bigcap_{R\in\mathcal{Q}(\phi)} \mathrm{EW}(R) \\ &\subseteq \bigcap_{R\in\mathcal{R}_0} \mathrm{EW}(R) = \{p\}.\end{aligned}$$

This implies that ϕ is superficially local so the first condition for $p \in \mathrm{Loc}(M)$ is satisfied. If it happens that the operator ϕ' above satisfies $[\phi'] = [\phi]$, our argument still applies and we must therefore have $\operatorname{supp}\phi' \subseteq \{p\}$. It is not possible to have $\operatorname{supp}\phi' = \emptyset$ since this would require that $[\phi'] = [\phi] = \mathcal{R}$ which is false. We conclude that $\operatorname{supp}\phi' = \{p\}$ and thus that $p \in \mathrm{Loc}(M)$.

We now prove the converse. Let p lie in $\mathrm{Loc}(M)$. Suppose that there does not exist any $\mathcal{R}_0 \subseteq \mathcal{R}$ with $\bigcap_{R\in\mathcal{R}_0} \mathrm{EW}(R) = \{p\}$. Let ϕ be a local operator at p which requires that ϕ is superficially local. There must exist a point $q \in M$ with

$$q \in \left(\bigcap_{R\in\mathcal{Q}(\phi)} \mathrm{EW}(R)\right) \setminus \{p\}.$$

Now consider a local operator ϕ' at the point q. Since q lies in the entanglement wedge of every region whose entanglement wedge contains p, we have $[\phi] \leq [\phi']$. The superficial locality of ϕ, along with the fact that ϕ' is not trivial, implies now that $[\phi] = [\phi']$ which, by the definition of $\mathrm{Loc}(M)$, implies that $\operatorname{supp}\phi' = \{p\}$ which is a contradiction. □

Theorem 5 is a useful tool for identifying examples of localizable regions in asymptotically AdS spacetimes as we will do in section 7.3. For now, we only advertise some facts that may be of interest. Localizable regions can extend quite far into the bulk spacetime. For the same reason that extremal surfaces can penetrate event horizons in some cases, $\mathrm{Loc}(M)$ can intersect a black hole interior. However, points that are too close to spacelike singularities are not localizable. Another interesting property of localizable regions is that they are not always subsets of the portion of the bulk that is accessible to boundary-anchored extremal codimension 2 surfaces with minimal area. In other words, $\mathrm{Loc}(M)$ can have a nonempty intersection with the entanglement shadow [16]. Before discussing these examples, however, we are going to introduce an object that will greatly increase the motivation for studying the localizable region.

The Space of Classes

The object that we now study is the collection of all equivalence classes of superficially local operators on G. We suggestively denote this set by $\tilde{M}$:

$$\tilde{M} = \left\{ [\phi] \,\middle|\, \phi \text{ is a superficially local operator on } G \right\}.$$

Given that an element $P \in \tilde{M}$ is a set of operators, all of which have the same value of $\mathcal{Q}$, it is convenient to let define $\mathcal{Q}(P)$ as $\mathcal{Q}(\phi)$ for any choice of $\phi \in P$.

An intuitive picture of $\tilde{M}$ is clear when $M = \mathrm{Loc}(M)$. In this case, Proposition 7.3.2 shows that there is a one-to-one correspondence between $\tilde{M}$ and M. The correspondence is that a point $p \in M$ is identified with the collection of all local operators at p. This reveals a new approach to bulk reconstruction from the boundary theory, somewhat similar in spirit to that of [61], which we will explore below.

Let us now make no assumptions about $\mathrm{Loc}(M)$ and determine the general structure of $\tilde{M}$. What we are going to find is that $\tilde{M}$ is equal to $\mathrm{Loc}(M)$ with the possible addition of some extra points in $\tilde{M}$. We refer to these unwanted extra points as "clumps."

First suppose that $p \in \mathrm{Loc}(M)$ and let ϕ be a local bulk operator at p. Then, $[\phi]$ consists only of local operators at p. (This follows directly from the definition of the localizable region.) As a consequence, a copy of $\mathrm{Loc}(M)$ can always be identified in $\tilde{M}$. Another thing that we can immediately show is that if Φ is any superficially local operator whose support consists of more than one point, then $\mathrm{supp}\,\Phi \cap \mathrm{Loc}(M) = \emptyset$. To see, this, suppose that $p \in \mathrm{supp}\,\Phi \cap \mathrm{Loc}(M)$ and consider a local operator ϕ at p. We would then have $[\Phi] \leq [\phi]$ with Φ superficially local so $[\Phi] = [\phi]$. This contradicts the definition of $\mathrm{Loc}(M)$ since Φ is nonlocal.

We cannot exclude the possibility that there exist nonlocal superficially local operators. To investigate this issue carefully, we introduce a map C that sends a point P in $\tilde{M}$ to a subset of M as follows:

$$C(P) = \bigcup_{\Phi \in P} \mathrm{supp}\,\Phi.$$

C has some nice properties:

Proposition 7.3.3. *Suppose that P and Q are elements of $\tilde{M}$. Then,*

1. *If every element of P is a local bulk operator, then there exists a point $p \in \mathrm{Loc}(M)$ such that $C(P) = \{p\}$,*
2. *if P contains a nonlocal operator, then $C(P) \cap \mathrm{Loc}(M) = \emptyset$,*
3. *if $C(P) \cap C(Q) \neq \emptyset$, then $P = Q$ and, in particular, C is injective.*

Proof. 1. If P consists of only local operators, then all of those operators must be at the same bulk point. To see this, suppose that ϕ_1 and ϕ_2 are two local bulk operators at bulk points p_1 and p_2 respectively. Now $\mathcal{Q}(\phi_1) = \mathcal{Q}(\phi_2)$ so any linear combination $\alpha\phi_1 + \beta\phi_2$ must

satisfy $[\phi_1] \leq [\alpha\phi_1 + \beta\phi_2]$. The superficial locality of ϕ_1 now proves that $[\phi_1] = [\alpha\phi_1 + \beta\phi_2]$ which contradicts the assumption that P consists only of local operators unless $p_1 = p_2$. Now let p denote the unique point in M where the elements of P are supported. It is obvious now that $C(P) = \{p\}$. Moreover, every local operator at p must lie in P and since there are no operators in P with support beyond $\{p\}$ we conclude that $p \in \mathrm{Loc}(M)$.

2. Assume that P contains a nonlocal bulk operator and suppose that $q \in C(P) \cap \mathrm{Loc}(M)$. Let ϕ denote a local operator at q. There must be some operator $\Phi \in P$ with $q \in \operatorname{supp} \Phi$ so $[\Phi] \leq [\phi]$ from which the superficial locality of Φ implies that $[\Phi] = [\phi]$ which is equivalent to the statement that $[\phi] \in P$. But this means that ϕ, a local operator in $\mathrm{Loc}(M)$, is equivalent to a nonlocal operator. This is a contradiction.

3. Suppose that there exists a bulk point $x \in C(P) \cap C(Q)$. Let ϕ denote a local operator at x. An argument identical to what was given for the proof of statement 2 shows that $\phi \in P$ and $\phi \in Q$. But P and Q are equivalence classes so the fact that they share an element means that $P = Q$. □

This argument shows that $\tilde{M}$ can be thought of as the union of $\mathrm{Loc}(M)$ with some extra points. Each extra point P has the property that $C(P)$ is a subset of M with more than one element. These objects are subtle enough to deserve a name:

Definition 4. *Suppose that $P \in \tilde{M}$ has the property that $C(P)$ has more than one element. Then, we will call both P and $C(P)$ a* clump.

Clumps are somewhat problematic because both local and nonlocal operators in clumps are superficially local. They therefore represent a potential threat to our approach. However, there is good news: we will argue in section 7.3 that clumps can be identified and removed using only the boundary theory (e.g. without relying on concepts like the bulk support of operators). Roughly speaking, clumps are associated with phase transitions for holographic entanglement entropy, and such phase transitions are visible in the boundary.

We are now in a position to give a much stronger answer to the fundamental question posed at the beginning of this section about identifying the operators on G that are dual to local operators in the bulk.

Theorem 6. *If there are no clumps, an operator ϕ on the code subspace G is dual to a local bulk operator in the localizable region if and only if ϕ is superficially local.*

If we assume the clump conjecture of section 7.3, which provides a way to identify and eliminate clumps, this conclusion provides the boundary dual to the concept of a bulk local operator (within a certain region of the bulk).

Examples

Examples can greatly clarify the machinery we have been developing. In particular, the spacetimes below demonstrate several features:

- Despite being associated with HRT surfaces, $\mathrm{Loc}(M)$ can probe entanglement shadows.
- $\mathrm{Loc}(M)$ can intersect black hole interiors (but it does not extend arbitrarily closely to spacelike singularities).
- In regions that are close to spacelike singularities, local operators are not superficially local.
- Clumps can occur, but the only known examples are associated with phase transitions where extremal surfaces "jump" around them.

Vacuum AdS

The simplest example is when M is vacuum AdS space (or any small perturbation of vacuum AdS) with dimension $D \geq 2+1$. For any point $p \in M$, theorem 5 immediately shows that $p \in \mathrm{Loc}(M)$. This is because in AdS space, one can always construct $D-1$ codimension 2 stationary surfaces intersecting p, whose tangent spaces at p are pairwise distinct, and then find the corresponding boundary regions $R_1, \ldots R_{D-1}$ on which these stationary surfaces are anchored. To prove that $p \in \mathrm{Loc}(M)$, we then consider the collection of regions $\{R_1, \ldots R_{D-1}, \bar{R}_1, \ldots, \bar{R}_{D-1}\}$ and apply this set to theorem 5.

Conclusion: If we somehow know that G is dual to a spacetime close to vacuum AdS, then an operator on G is local if and only if it is superficially local. The space of classes of superficially local operators, $\tilde{M}$, is a reconstruction of the bulk.

Conical AdS

Anti-de Sitter space with a conical deficit is a simple example of a spacetime with an entanglement shadow[7] [16]. Given that $\mathrm{Loc}(M)$ can be defined by means of HRT surfaces, one might suspect that for conical AdS, $\mathrm{Loc}(M)$ is a proper subset of M. We will explain why this is not the case and that, in fact, we again have $\mathrm{Loc}(M) = M$.

Let n be an integer greater than 1 and consider, for example, $M = \mathrm{AdS}_{2+1}/\mathbf{Z}_n$. The metric can be written as

$$ds^2 = -\left(\frac{1}{n^2} + \frac{r^2}{L^2}\right) dt^2 + \left(\frac{1}{n^2} + \frac{r^2}{L^2}\right)^{-1} dr^2 + r^2 d\phi^2 \tag{7.9}$$

where $-\infty < t < \infty$, $r > 0$, and $\phi \in [0, 2\pi)$. There is a critical radius r_{crit} such that no HRT surface intersects the region $r < r_{\mathrm{crit}}$. If $\{R_s\}$ is a continuous nested family of boundary regions with R_{-1} a small region and R_1 wrapping around almost the entire boundary, the HRT surface anchored to R_s, $\mathrm{ext}\, R_s$, will discontinuously jump around the shadow at some

[7] To our knowledge, [16] and related work has only studied regions that are not probed by minimal surfaces anchored to static boundary regions rather than the general stationary surfaces appearing in the calculation of covariant holographic entanglement entropy. Below we assume that the general features of the entanglement shadow in standard examples are unchanged if non-static surfaces are considered.

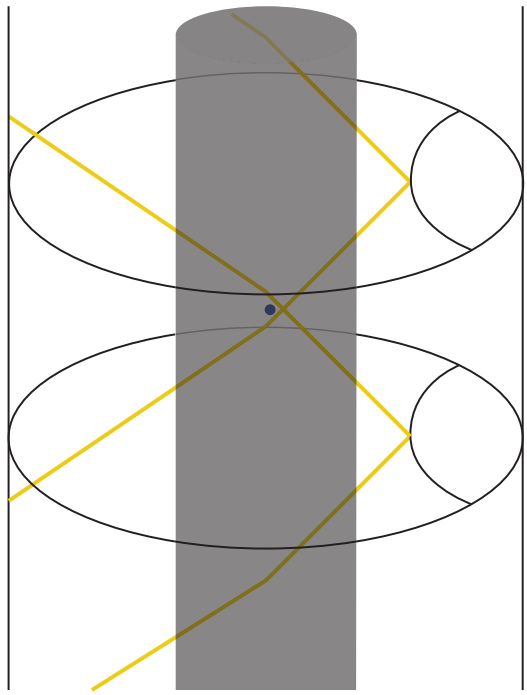


Figure 7.3: Conical AdS is an example of how points in the bulk that are not directly probed by extremal surfaces can still be in the localizable region. Despite the entanglement shadow (the grey cylinder), points can be localized because they can intersect boundaries of entanglement wedges.

critical value of s. Note that this phenomenon is not related to extremal surface barriers [62] but is instead a consequence of there being more than one stationary codimension 2 surface anchored to any given boundary region: no HRT surface enters the shadow because there would always be another stationary surface that does not enter the shadow with smaller area. The discontinuous jump can be regarded as a phase transition in the sense that the von Neumann entropy $S(R_s)$, regarded as a function of the parameter s, has a discontinuous derivative at the jump.

If $p \in M$ lies outside of the entanglement shadow, we must have $p \in \mathrm{Loc}(M)$ for the same reason that every point is localized in vacuum AdS. On the other hand, suppose that p lies within the entanglement shadow. To show that $p \in \mathrm{Loc}(M)$, all we need, by theorem 5, is a finite set of boundary regions such that the intersection of their entanglement wedges is $\{p\}$.

This can by done by considering regions like those shown in figure 7.3. Note that only two regions are shown in the figure but that the point can be completely localized by adding other boundary regions such as rotations of the regions depicted. The trick here is easy to understand: it is not necessary for HRT surfaces to intersect localized points as long as boundaries of entanglement wedges intersect them instead.

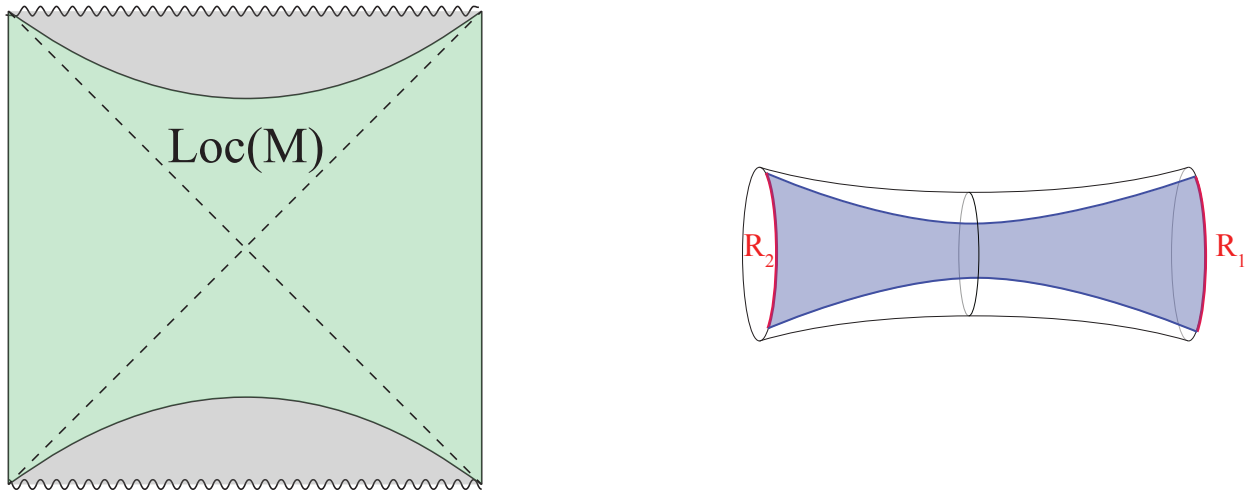


Figure 7.4

Conclusion: If G is dual to a spacetime close to $\mathrm{AdS}_{2+1}/\mathbf{Z}_n$, then an operator on G is local if and only if it is superficially local. The space of classes of superficially local operators, $\tilde{M}$, is a reconstruction of the bulk.

Two-Sided Black Holes

In the case where M is an eternal AdS-Schwarzschild geometry, which has two disconnected boundary components, the localizable region extends into the black hole interior but does not probe all the way to the singularity. This is depicted in figure 7.4. Many points in the interior region can be localized by considering boundary regions that consist of two disconnected components lying in different boundaries (see figure 7.4). HRT surfaces, however, do not reach points that are arbitrarily close the future or past singularities: there is a critical radius r_{crit} (smaller than the black hole radius) that no boundary-anchored extremal surface extends beyond [187, 62]. Figure 7.5 proves that local operators at points with radius $r < r_{\text{crit}}$ are not superficially local. This portion of the spacetime is completely missed by our methods and will thus be called the *inaccessible region*.

Conclusion: If G is dual to an eternal AdS-Schwarzschild geometry (with two boundary CFTs), then an operator ϕ on G is superficially local if and only if it is dual to a local bulk operator at a bulk point with $r > r_{\text{crit}}$. The space of classes of superficially local operators, $\tilde{M}$, is a reconstruction of the region of M with $r > r_{\text{crit}}$.

Dynamical Black Holes

The previous example might have given the impression that $\mathrm{Loc}(M)$ cannot intersect a black hole interior without appealing to entanglement between two CFTs. This is not the case.

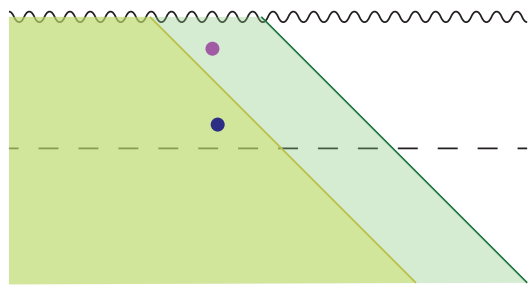


Figure 7.5: When a point (purple) is close to a spacelike singularity, it is very difficult for the point to be in Loc(M). Quite generally, HRT surfaces are prevented from approaching such singularities [187, 62]. In this figure, the horizontal dashed line is a surface with the property that no HRT surface intersects its future. (This is more restrictive than an extremal surface barrier, which would prohibit smooth deformations of stationary surfaces.) A local operator at the purple point cannot be superficially local since a point in its past (blue) will typically be contained in strictly more entanglement wedges.

Consider a black hole that forms from collapse in an asymptotically AdS spacetime. Then, it has been demonstrated [97] that HRT surfaces probe the black hole interior (although they do not approach the singularity arbitrarily closely). Because such HRT surfaces can be anchored to boundary regions at a variety of angular positions, we conclude that Loc(M) enters the black hole interior in this case. Note, however, that figure 7.5 again explains why regions too close to the singularity are not localizable.

Bag of Gold

Our fourth example is a "bag of gold" spacetime (see, e.g., [125]). The manifold M is an AdS-Schwarzschild spacetime with one of its two asymptotic regions removed and replaced with a patch of de Sitter space. The spacetime is static and spherically symmetric. Its Penrose diagram is shown in figure 7.6. We will label the regions in the diagram I-IV as shown in the figure (note that region II includes the de Sitter patch). It is very important to understand that unlike the two-side AdS-Schwarzschild spacetime, M has only one asymptotic boundary with topology $S^{D-2} \times \mathbf{R}$. The time slice Σ that is marked in figure 7.6 has the topology of $\mathbf{R^{D-1}}$. In particular, Σ is simply connected and the homology constraint for HRT surfaces will not play any interesting role here. The dotted line in region I is a surface beyond which no HRT surface probes.

We will argue the following.

1. Loc(M) is the portion of region I that is probed by HRT surfaces.

2. $\tilde{M}$ has a single clump whose image under C (see section 7.3) is all of region II. Thus, we will say that region II is a clump.

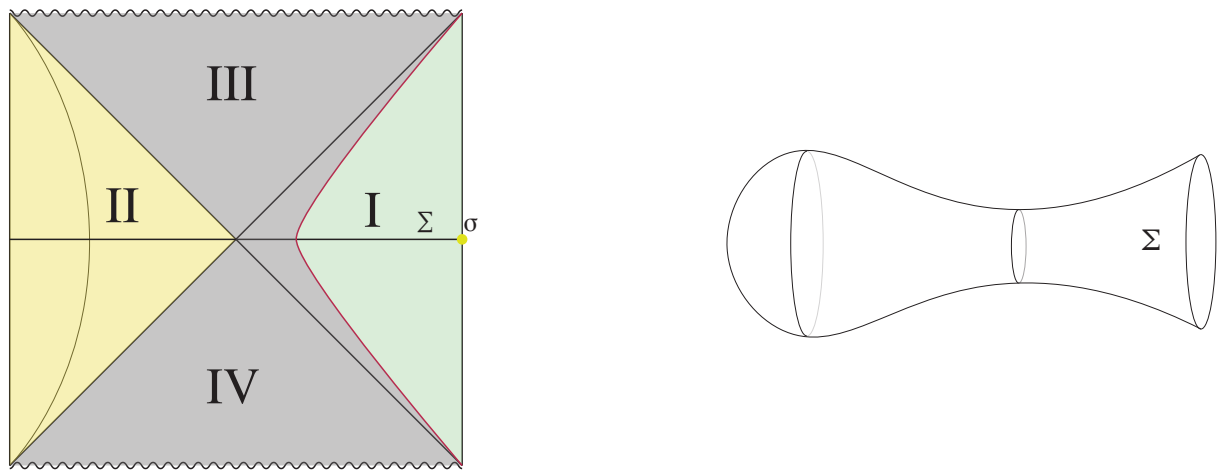


Figure 7.6: The bag of gold geometry we consider is obtained by removing an asymptotic region from an AdS black hole and replacing it with a patch of de Sitter space. As discussed in the text, the localizable region is the portion of region I that is accessible to HRT surfaces and region II is a single clump. The remaining portion of the spacetime is "inaccessible" in the sense that no operator with support in these regions is superficially local.

3. The rest of the spacetime (including regions II and IV) is neither localizable nor within clumps. It is "inaccessible."

First let us discuss why region II is a clump. Like conical AdS, this spacetime exhibits phase transitions in its HRT surfaces as well as an entanglement shadow. Consider the boundary time slice $\sigma = \partial\Sigma$ and let R_ψ be a spherical cap on σ with opening angle ψ (defined so that $R_\pi = \sigma$). The spacetime in region I is identical to region I of AdS-Schwarzschild so the structure of stationary codimension 2 boundary-anchored surfaces must also be the same and, in particular, there are always two distinct stationary surfaces anchored to R_ψ. At $\psi = \pi/2$, there is a phase transition with a discontinuity in the first derivative of $S(R_\psi)$. At this transition, the minimal surface jumps around the entire region II. Note also that HRT surfaces fail to even contact the bifurcation throat: there is, once again, a minimal radius in region I, r_{crit}, greater than the black hole radius, within which no HRT surface extends.

If $\psi < \pi/2$, the spatial region V_ψ on Σ between R_ψ and its HRT surface ext R_ψ is confined to region I. Thus $\text{EW}(R_\psi)$ is confined to region I; this follows from the fact that $\text{EW}(R_\psi) = D(V_\psi)$ after compactification. Meanwhile, When $\psi > \pi/2$, V_ψ contains the entire intersection of Σ with region II and $\text{EW}(R_\psi)$ must contain all of region II. These observations were made for a simple spherical cap on the time-reversal symmetric slice Σ, but they hold very generally: any time we consider a nested family of boundary regions $\{R_s \in \mathcal{R}\}$, $\text{EW}(R_s)$ is confined to region I for s smaller than some critical value and $\text{EW}(R_s)$ contains all of region II when s exceeds this value.

What this shows is that if ϕ_1 and ϕ_2 are two bulk operators with support in region II, we must have $\mathcal{Q}(\phi_1) = \mathcal{Q}(\phi_2)$. Moreover, note that any operator ϕ which is supported in region II must be superficially local. To see this, consider any $x \in M \setminus (\text{ region II})$. If x is in region I or III, take a spherical cap like R_ψ with $\psi > \pi/2$, but place it on a boundary time slice at very early time. No matter how early time time is taken, time-translation invariance guarantees that region II is still contained in $\mathrm{EW}(R_\psi)$, but by sending the boundary time slice to $-\infty$, we can put any point in regions I or III in the future of ext R_ψ. This means that there exists some $R \in \mathcal{Q}(\phi) \setminus \mathcal{Q}(\phi')$ so $\phi \not\leq \phi'$. The same argument can be made if x is in region IV by sending the boundary time slice to $+\infty$. We conclude that ϕ must be superficially local and region II is thus a clump (since all operators with support in region II are superficially local and have the same image under $\mathcal{Q}$).

Let us finally study the remainder of the spacetime. The portion of region I that is probed by HRT surfaces is readily seen to be contained in $\mathrm{Loc}(M)$. We now outline an argument that, in fact, this probed region is exactly $\mathrm{Loc}(M)$. Figure 7.5 gives an explanation of why local operators in region III cannot be localized. More generally, consider a local bulk operator ϕ_x at a point x lying outside of the region probed by HRT surfaces but also lying outside of the clumped region II. If R is a boundary region with $x \in \mathrm{EW}(R)$, then R must be large enough to have undergone a phase transition so that region II is contained in the entanglement wedge of R as well. This means that if Φ is any superficially local operator in the clump, we have $\mathcal{Q}(\phi_x) \subsetneq \mathcal{Q}(\Phi)$. This shows that ϕ_x cannot be superficially local.

Conclusion: Suppose that G is dual to the bag of gold geometry. If ϕ is a superficially local operator, then it is either a local operator in the portion of region I probed by extremal surfaces or it is some operator (which need not be local) with support in region II. The clump conjecture of section 7.3 is valid for this spacetime, so the problematic superficially local operators can be identified and discarded. After doing so, the remaining superficially local operators exactly form the collection of all bulk local operators in $\mathrm{Loc}(M)$.

The Clump Conjecture

In this section we propose a way to use the boundary theory to identify and remove clumps from $\tilde{M}$. Specifically we give an alternative definition of a clump that does not make direct reference to the bulk and we conjecture that our two definitions are equivalent. We know of no counterexamples to the conjecture and there is good evidence for its general validity.

The basic motivation is as follows. If $P \in \tilde{M}$ is a clump, then, by definition, $C(P)$ contains more than one bulk point. Generically, clumps have nonzero spacetime volume. On the other hand, we know that no entanglement wedge can contain only part of a clump: if $R \in \mathcal{R}$, then either $C(P) \subseteq \mathrm{EW}(R)$ or $C(P) \cap (\mathrm{EW}(R))^\circ = \emptyset$. These observations indicate that if R_s is a continuous nested one-parameter family of regions in $\mathcal{R}$ such that $R_s \in \mathcal{Q}(P)$ for $s > 0$ and $R_s \notin \mathcal{Q}(P)$ when $s < 0$, we must have some form of a discontinuity in the entanglement wedges $\mathrm{EW}(R_s)$ as a function of s at $s = 0$. Such discontinuities occur when the HRT surfaces anchored to $\{R_s\}$ jump discontinuously. But such a jump can often be seen

in the boundary theory in the form of a discontinuity in a derivative of the von Neumann entropy of the boundary regions R_s.

Before stating the conjecture formally, we give a useful definition:

Definition 5. *Let ϕ be an operator on G and $R \in \mathcal{Q}(\phi)$. R is said to be* minimal *if whenever $R' \subsetneq R$, $R \notin \mathcal{Q}(\phi)$.*

We will also introduce the map $\bar{\mathcal{Q}}$ by letting $\bar{\mathcal{Q}}(\phi)$ denote the collection of minimal elements of $\mathcal{Q}(\phi)$. Additionally, if $P \in \tilde{M}$, we will define $\bar{\mathcal{Q}}(P)$ as $\bar{\mathcal{Q}}(\phi)$ for any choice of $\phi \in P$ (all choices of ϕ have the same $\bar{\mathcal{Q}}(\phi)$).

As suggested above, phase transitions in the boundary theory will play a role in the boundary identification of clumps. To be clear, a "phase transition" refers to the following situation. Suppose that $\{R_s \big| -1 < s < 1\}$ is a regular[8] one-parameter family of boundary regions with $R_{s_1} \subsetneq R_{s_2}$ whenever $s_1 < s_2$. Let $S(R_s)$ denote the von Neumann entropy of the boundary region R_s in any state[9] in the code subspace G. We say that there is a phase transition at $s = 0$ if some derivative of $S(R_s)$ at $s = 0$ is discontinuous. Moreover, if $R \in \mathcal{R}$, we will say that there is a phase transition at R if there is some one parameter deformation of the form above, $\{R_s\}$, with $R_0 = R$.

We now state our proposal for identifying and removing clumps. We will refer to it as the *clump conjecture*:

Suppose that $P \in \tilde{M}$. P is a clump if and only if for every $R \in \bar{\mathcal{Q}}(P)$, there is a phase transition at R.

We immediately note that this conjecture is consistent with the examples provided in section 7.3. The only example we gave of a clump is that of the bag of gold spacetime which always features phase transitions for minimal regions. Consider, however the example of $\mathrm{AdS}_{2+1}/\mathbf{Z}_n$. This may appear to contradict the clump conjecture because it is a spacetime with no clumps but which does posses phase transitions. However, consider regions like the ones depicted in figure 7.3. These are indeed minimal regions for the operator at the point depicted (which corresponds to a point in $\tilde{M}$. However, there is no phase transition at such a region. This is why the statement of the clump conjecture requires that there is a phase transition for *every* $R \in \bar{\mathcal{Q}}(P)$.

[8] By "regular" we mean that R_s deforms smoothly enough that we are not introducing discontinuities in any derivative of von Neumann entropy by choosing an awkward parameterization of regions.

[9] $S(R_s)$ is state-dependent, but the spacetime background is approximately fixed within the code subspace G, so assertions about phase transitions will be state-independent at leading order.

7.4 Reconstruction of Causal Structure and Beyond

From here on we assume the validity of the clump conjecture (which we strongly expect) and use a new definition of $\tilde{M}$:

$$\tilde{M} = \big\{[\phi] \,\big|\, \phi \text{ is superficially local and } [\phi] \text{ is not a clump}\big\}$$

This can be done using only the boundary theory. Simply begin with $\tilde{M}$ as defined previously, and then remove clumps from it by using the clump conjecture.

With this new definition, a major conclusion of section 7.3 is that in some sense $\tilde{M}$ is isomorphic to $\mathrm{Loc}(M)$ although we have not been very clear about what sort of isomorphism this is. We are now going to take the view that $\tilde{M}$ can be thought of as a reconstruction of the bulk very seriously. We will successfully determine a metric on $\tilde{M}$ up to a conformal rescaling. This will be done using only information available in the boundary theory (which includes the definition of $\tilde{M}$ itself). The manifold $\tilde{M}$ and its causal structure will exactly reproduce that of $\mathrm{Loc}(M)$. This constitutes a boundary reconstruction of the metric on $\mathrm{Loc}(M)$ up to its conformal factor.

Spacelike Separation and Microcausality

The key insight to identifying a causal structure on $\tilde{M}$ is to note that $\tilde{M}$ consists of collections of operators on the code subspace G and that the commutation relations amongst those operators must betray an aspect of the bulk spacetime geometry. This suggests the following definition:

Definition 6. *Suppose that $P, Q \in \tilde{M}$. We say that P and Q are* spacelike separated *if for every $\phi_1 \in P$ and $\phi_2 \in Q$, we have $[\phi_1, \phi_2] = 0$. Otherwise, we say that P and Q are* causally related.

There are two things to immediately notice about this definition. First, while we have defined the statement that P and Q are causally related, we have not yet given meaning to the statement that P is to the future of Q. This will be addressed below. Second, note that for P and Q to be causally related, all that is necessary is that there exists some $\phi_1 \in P$ and some $\phi_2 \in Q$ such that ϕ_1 and ϕ_2 fail to commute. It is certainly not necessary that all such operators would fail to commute.

In special cases, it is possible to conclude that P and Q are spacelike separated without relying directly studying the commutativity of their operators. If it happens that there exists $R_1 \in \mathcal{Q}(P), R_2 \in \mathcal{Q}(Q)$ with the property that R_1 and R_2 are spacelike separated in the boundary, meaning that

$$\big(J_+^\partial(R_1) \cup J_-^\partial(R_1)\big) \cap R_2 = \emptyset,$$

then microcausality in the boundary field theory guarantees that any operators O_1 and O_2 in the algebras of R_1 and R_2 respectively must have $[O_1, O_2] = 0$. In particular, for any

$\phi_1 \in P$ and $\phi_2 \in Q$, we can find reconstructions of ϕ_1 and ϕ_2 in R_1 and R_2 respectively and conclude that $[\phi_1, \phi_2] = 0$. However, this situation is too much to ask for in general.

In the case where two classes P and Q are causally related, the above logic indicates that there absolutely cannot be any $R_1 \in \mathcal{Q}(P), R_2 \in \mathcal{Q}(Q)$ with the property that R_1 and R_2 are spacelike separated in the boundary. This is consistent with a theorem in bulk geometry which is a necessary result for the consistency of entanglement wedge reconstruction:

Proposition 7.4.1. *Let M be an asymptotically AdS spacetime and suppose that $p, q \in M$ are bulk points with $q \in I_+(p)$. Suppose, moreover, that there exist boundary regions $R_1, R_2 \in \mathcal{R}$ such that $p \in \mathrm{EW}(R_1), q \in \mathrm{EW}(R_2)$. Then, $\left(I_+^\partial(R_1) \cup I_-^\partial(R_1)\right) \cap R_2 \neq \emptyset$.*

Proof. Choose a Cauchy surface σ of ∂M with $R_1 \subseteq \sigma$ and let $\bar{R}_1 = \sigma \setminus R_1$. Let Σ be any AdS-Cauchy surface for the bulk with $\partial\Sigma = \sigma$ and write $\Sigma = S \cup \bar{S}$ where $S \cap \bar{S}$ is the HRT surface of R_1. Then, $q \notin \mathrm{EW}(\bar{R}_1)$. (This follows from the fact that $\mathrm{EW}(R_1) = D(S)$ and $\mathrm{EW}(\bar{R}_1) = D(\bar{S})$.)

Suppose that we had $R_2 \subseteq D^\partial(\bar{R}_1)$. Wall's entanglement wedge nesting theorem [187] implies that this would require that $\mathrm{EW}(R_2) \subseteq \mathrm{EW}(\bar{R}_1)$ which contradicts the fact that $q \in \mathrm{EW}(R_2)$. Thus, R_2 is not contained (entirely) in $D^\partial(\bar{R}_1)$. On the other hand, the boundary is flat so $D(\bar{R}_1) = \partial M \setminus (I_+^\partial(R_1) \cup I_-^\partial(R_1))$. We conclude that R_2 intersects $I_+^\partial(R_1) \cup I_-^\partial(R_1)$. □

Time Orientation

Suppose that P and Q are points in $\tilde{M}$ that are causally related. Then, the corresponding bulk points, p and q respectively, must either have $p \in J_+(q)$ or $q \in J_+(p)$. But how do we know which?

There may be a very direct way to answer this question. Here, however, we give a topological answer. In appendix .1 we explain how $\tilde{M}$ be be made into a topological space. The basic idea is fairly obvious: two points in $\tilde{M}$ are close to each other if their images under $\mathcal{Q}$ are close. Because this topology will be consistent with the bulk topology on $\mathrm{Loc}(M)$, we can make use of topological features of the causal structure of the spacetime $\mathrm{Loc}(M)$.

Of particular use is the fact that if $p \in M$, $J_+(p)$ is connected (as is $J_-(p)$). Because $\mathrm{Loc}(M)$ may be a proper subset of M, it is possible that $J_+(p) \cap \mathrm{Loc}(M)$ is not connected. Nonetheless, we can consider the connected component of $J_+(p) \cap \mathrm{Loc}(M)$ that contains p. The same construction must be possible in $\tilde{M}$, but we have to be somewhat more careful. For $P \in \tilde{M}$, we can consider the set of points K that are causally related to P. This includes P itself. We can then consider $K \setminus \{P\}$ and look at the two connected components of K that are arbitrarily close to P. (There must be exactly two such components because the topology on $\tilde{M}$ needs to be consistent with that of $\mathrm{Loc}(M)$.) We label these two components $\tilde{J}_\pm(P)$ with the understanding that we have yet to determine which component deserves a plus sign and which deserves a minus sign.

Suppose we arbitrarily choose which of the two regions is to be called $\tilde{J}_+(P_0)$ for one particular point P_0. In all but the most pathological of connected spacetimes, this fixes the

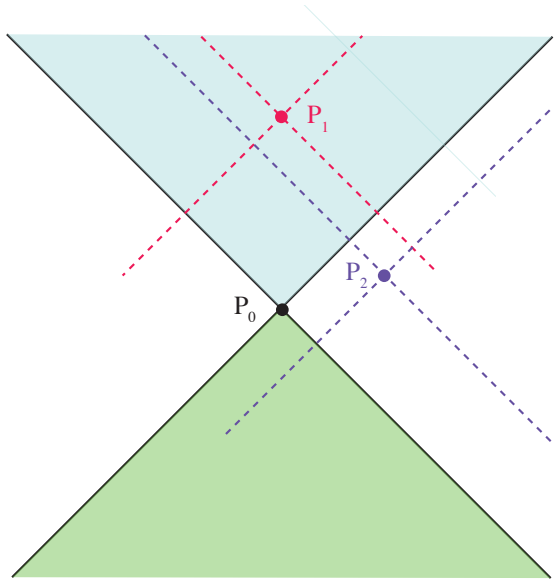


Figure 7.7: If the definition of the future and past of a point $P_0 \in \tilde{M}$ is chosen, there is an immediate constraint on the time orientation at other points in $\tilde{M}$. In this figure, the orientation at P_0 also fixes the orientation at P_1 and P_2.

time orientation for every other point in the spacetime. For example, suppose that P_1 is another point in $\tilde{M}$ and that $P_1 \in \tilde{J}_+(P_0)$. (Here we are making use of our arbitrary decision about $\tilde{J}_+(P_0)$.) Then, we must assign the orientation at P_1 so that $P_0 \in \tilde{J}_-(P_1)$. But now, if we find another point $P_2 \in \tilde{J}_-(P_1)$, we must have that $\tilde{J}_+(P_2)$ contains P_1. Continuing in this way, we can expect to be able to fix the time orientation for every point in $\tilde{M}$ as long as it is connected. This process is depicted in figure 7.7

But what about the overall time orientation? That is, how do we decide on $\tilde{J}_+(P_0)$ in our example above? This can be done by beginning with a point in $\tilde{M}$ that corresponds to local boundary operators at some boundary point. On ∂M, we already have a notion of future and past. Thus, if we take P_0 to be an equivalence class consisting only of local boundary operators at a point $x \in \partial M$, we can decide upon $\tilde{J}_+(P_0)$ by requiring that if P_1 is another class of local boundary operators lying at a point y then $P_1 \in \tilde{J}_+(P_0)$ only if $y \in J_+^{\partial}(x)$.

We have now succeeded in defining a causal structure on $\tilde{M}$ that must be consistent with that on $\mathrm{Loc}(M)$. As a consequence, we have reconstructed the metric in $\mathrm{Loc}(M)$ up to an undetermined conformal factor.

Comparison with Light-Cone Cut Reconstruction

There is a compelling connection between the bulk reconstruction developed here and a recent approach to bulk reconstruction involving light-cone cuts due to Engelhardt and Horowitz

[61]. Cut reconstruction is a new area of research [59, 57, 60, 58], and remarkably, a number of the ideas involving cuts appear to have analogs in superficial locality reconstruction. We now detail the similarities and differences between the two approaches.

- *Large N*: Both cut reconstruction and reconstruction with superficial locality require in their current forms that the classical limit be taken. Light cone cuts are associated with singularities of correlation functions of local boundary operators that only resolve in the large N limit. These singularities in the boundary theory are at first mysterious but have a simple explanation if one knows about the dual bulk: if there is a bulk point p in the causal wedge of the boundary, then cut singularities are singularities of boundary n-point functions $< O(x_1) \dots O(x_n) >$ that can occur when the boundary points lie on the *future and past cuts* of p: $C_\pm(p) = (\partial J_\pm(p)) \cap \partial M$. These singularities are generally known as bulk-point singularities and have been considered in several contexts prior to that of cuts [69, 123]. In particular, [123] provided an example showing that such singularities are not expected to arise without sending N to infinity. This is consistent with the fact that there should not be any notion of a local bulk scattering point when N is finite. Similarly, our consideration of superficially local operators and their equivalence classes is certainly only expected to reproduce local bulk physics in the large N limit. At finite N there are no local (gauge-invariant) observables in the quantum gravity [51, 52], so it is not clear why one would even seek to study any notion of exactly local bulk operators in this case. It is, of course, interesting to contemplate whether or not either of these approaches suggests new ways to think about approximate locality at finite but large N.

- *Specification of a state*: Cut reconstruction, in its original form, presupposes that we are given a particular quantum state ψ in the CFT Hilbert space and that we are told that ψ is dual to some *unknown* bulk geometry.[10] The task is then to study correlation functions in that state (which can be done using the boundary theory only) to determine aspects of the bulk interpretation of ψ (like the bulk geometry). Similarly, throughout this paper we have assumed that we are given a code subspace G and that we are told that G has the bulk interpretation of being the Hilbert space of a quantum field theory on some unknown spacetime background. We then consider various operators acting on G and ask which of them are superficially local (which can be done using the boundary theory only).

- *Identification of points with a boundary object*: The next step in cut reconstruction is to make an identification between the set of light-cone cuts and the set of points in the causal wedge of the boundary. On the other hand, here we identify points in $\mathrm{Loc}(M)$ with equivalence classes of superficially local operators.

[10] In [58], the theory of cuts was put into a framework that did not strictly rely on the presumption of the existence of a bulk, but where an extra dimension can be seen to emerge in appropriate cases.

- *Reconstruction of the Conformal Metric*: It is possible to assign a causal structure to the set of cuts. This causal structure is consistent with the causal structure in the set of bulk points corresponding to the cuts (with some caveats that can be addressed). As a result, the set of cuts provides a reconstruction of the metric in the causal wedge of the boundary, $\mathrm{CW}(\partial M)$, up to a conformal factor. Similarly, we are able to identify a causal structure on $\tilde{M}$, the set of classes of superficially local operators, and we therefore obtain a reconstruction of the conformal class of the metric in the bulk region $\mathrm{Loc}(M)$. It is known that in some cases, $\mathrm{Loc}(M)$ extends further into the bulk than $\mathrm{CW}(\partial M)$ does: in the case of a dynamical black hole, $\mathrm{Loc}(M)$ can intersect the black hole interior. We do not know whether or not it is always the case that $\mathrm{CW}(\partial M) \subseteq \mathrm{Loc}(M)$.

- *Local operators and the connection between the two methods*: The premise of our approach was to solve a different problem from bulk reconstruction. Superficial locality provides a way to identify the operators on a code subspace G that are dual to local bulk operators. Identification of bulk local operators has not yet been a goal of light-cone cut reconstruction, but it is a promising direction. In fact, such considerations suggest a way to directly relate cut reconstruction to our program. Consider a point $P \in \tilde{M}$ and also consider a light-cone cut $C_\pm$ associated with singularities in correlation functions computed in a state $\psi \in G$. We would like to know how to tell if the bulk point associated with P is the same as the bulk point associated with $C_\pm$ (clearly this is only plausible for bulk points in the intersection of $\mathrm{CW}(\partial M)$ and $\mathrm{Loc}(M)$.

  We suggest the following approach to this problem. Consider a superficially local operator $\phi \in P$ and take a collection of boundary points $x_1, \ldots, x_n$ close to points in C. Now, consider two different correlation functions:

$$F_n(x_1, \ldots, x_n) = \langle \psi | O(x_1) \ldots O(x_n) | \psi \rangle$$
$$G_n(x_1, \ldots, x_n) = \langle \psi | \phi \, O(x_1) \ldots O(x_n) | \psi \rangle.$$

  If ϕ is indeed a local operator at the vertex of the cut C, then a signature of that property will be encoded in the relationships between F_n and G_m for various values of n and m. We do not pursue this idea further in the present work.

7.5 Discussion

Relying only on subregion duality between the boundary and bulk spacetimes, our construction addresses the following question. Given a CFT and a code subspace dual to an unknown geometry, can we tell if some operator is dual to a bulk local operator? To answer this question, we exploit the curious feature that numerous distinct boundary regions can reconstruct a local bulk operator. Once we identify the set of local bulk operators in the localizable region, the relations among those operators reveal bulk causal structure.

Furthermore, because the program focuses on entanglement wedges, as opposed to the extremal surfaces themselves, the operators we identify can lie behind horizons and within entanglement shadows in many examples. As expected, however, there are still regions for which our procedure fails to completely describe locality (these regions are often behind horizons). If we assume bulk locality still holds even within these regions, its encoding in the CFT is different than that of operators in the localizable region.

Subregion duality is a common property of holography. The holographic entanglement entropy prescription [157, 99] and the fact that entanglement wedge reconstruction is possible [49], lead us to the conclusion that quantum error correction is a feature of any theory with a holographic description. This is an extra constraint on holographic CFTs, which must encode information in a way consistent with bulk reconstruction, and can be seen as a requirement of CFTs having a bulk dual.

Remaining Considerations

Finite N: While we have addressed how locality, for the portion of the bulk in the localizable region, emerges from quantum error correction, there are still gaps that need to be understood. To what extent does locality fail at finite N? Gravitational effects prohibit the existence of local bulk observables. However, the quantum error correcting properties of subregion duality hold beyond leading order and it may therefore be elucidating to consider an approximate form of our approach at finite N. This may shed light on the subtleties of the large N limit and the relationship between exact quantum gravity and the infinite N theory.

The conformal factor: While there is no obvious way to reconstruct the conformal factor on $\tilde{M}$, we can argue that more information than just the causal structure is available to us. Consider a point $P \in \tilde{M}$ with the special property that for some boundary region $R \in \mathcal{R}$, both R and its complement $\bar{R}$ lie in $\mathcal{Q}(P)$. The only geometrical interpretation of this scenario is that operators in P correspond to a point on the HRT surface ext R. This means that in addition to the conformal metric on $\tilde{M}$ we also know the minimal area anchored extremal surfaces as well as the (regulated) areas of those surfaces, determined by the von Neumann entropies of corresponding boundary regions [157, 99, 113, 50]. Noting that stationary surfaces and their areas are not invariant under conformal transformations, the conformal factor on the metric is significantly constrained. We leave further investigation in this direction to future work.

.1 Appendix for Chapter 7

The large N limit

In discussing the main concepts in the text, we have assumed that local bulk operators exist, hoping to present our construction in an intuitive fashion. However, exact bulk locality only exists when $N = \infty$, and gravitational effects are turned off. When N is large but finite,

gravitational effects demands that any gauge invariant bulk operator will be nonlocal in some way [51, 52]. Nevertheless, bulk effective field theory still makes sense within the code subspace of holographic CFTs. This is possible because nonlocal effects become small, since they come with some positive power of the gravitational coupling. The suppression in N allows us to discuss local bulk fields (perhaps smeared over a region $\sim l_p$) and perturbatively add nonlocal effects (by appropriately dressing the fields for example), so long as we work in the appropriate code subspace.

Here, we explain how the constructions in the main text can be made precise by appropriately applying the large N limit to decouple nonlocalities due to gravity. Consider a CFT satisfying the appropriate requirements for having a bulk dual (see e.g. [95]). The theory has some parameter, $\epsilon(N)$, which corresponds to the gravitational coupling in the bulk and taking $\epsilon \to 0$ means turning off gravitational effects (i.e. sending $N \to \infty$). Different values of ϵ correspond to different boundary theories (with different central charges) with an associated Hilbert space H_ϵ.

For $\epsilon \neq 0$, no gauge-invariant operator ϕ_ϵ, restricted to the appropriate code subspace G_ϵ, will be local in the bulk. However, as we decrease ϵ, the strength of nonlocal gravitational effects decreases, and some operators and some operators in the CFT will start to resemble what one expects for local operators in semiclassical field theory; intuitively these would be the operators that would limit to local fields in the $\epsilon = 0$ limit. For example, if we think about semiclassical fields that are gravitationally dressed, the gravitational coupling suppresses the nonlocal dressing.

Consider now a family of operators, $\{\phi_\epsilon\}_{\epsilon>0}$, with ϕ_ϵ acting on the code subspace G_ϵ for all $\epsilon > 0$. [11]

Definition 7. *Let $R \in \mathcal{R}$ be a boundary region and let $\bar{R}$ be a complement of R. We say that a family $\{\phi_\epsilon\}_{\epsilon>0}$ is* reconstructable *in R if for any family of operators $\{O_\epsilon^{\bar{R}}\}_{\epsilon>0}$ in the algebra of $\bar{R}$ for H_ϵ and for any family of states $\{\psi_\epsilon\}_{\epsilon>0}$ with $\psi_\epsilon \in G_\epsilon$,*

$$\lim_{\epsilon\to 0}\langle\psi_\epsilon|[\phi_\epsilon, O_\epsilon^{\bar{R}}]|\psi_\epsilon\rangle = 0 \tag{10}$$

As reviewed in section 7.2, this implies that, when ϵ is very small, there is some operator O_ϵ^R in the algebra of R, whose action on G_ϵ is that of the operator ϕ_ϵ (up to corrections in ϵ).

Note that most of these families of operators will not limit to a semiclassical local bulk field. The "limit" might be a smeared operator in $\mathrm{EW}(R)$ or the family of operators could oscillate forever within $\mathrm{EW}(R)$ and never converge in any sense. However, some special class of such families do limit to local operators.

In order to test whether or not a collection of operators approaches a local field as ϵ becomes small, we introduce a generalization of the procedure in the text. The idea is to

[11] Decreasing ϵ decreases the strength of gravitational backreaction. In order to keep any nontrivial background fixed while changing the value of ϵ, we must separate "background matter" from excitations. As we send $\epsilon \to 0$, the stress tensor for the background matter must be rescaled appropriately to maintain a nontrivial background. This emphasizes the subtlety in the definition of G_ϵ

make the fundamental object of study the collection of ϵ-dependent families of operators as opposed to the set of operators on a fixed code subspace. Following the framework from section 7.3, we introduce a map $\mathcal{Q}$ that acts on families of operators as follows:

$$\mathcal{Q}(\{\phi_\epsilon\}_{\epsilon>0}) = \{R \,\big|\, \{\phi_\epsilon\}_{\epsilon>0} \text{ is reconstructible in } R\} \tag{11}$$

For some of these sequences, the set $\mathcal{Q}(\{\phi_\epsilon\}_{\epsilon>0})$ will be the result expected for a field localized to a point in the bulk. If this is the case, we can think of $\{\phi_\epsilon\}_{\epsilon>0}$ as a set of operators whose bulk interpretation is a semiclassical field (built on a background associated with a code subspace) whose nonlocal gravitational effects disappears as $\epsilon \to 0$. For such sequences of operators, taking the $\epsilon \to 0$ limit is can be thought of as "undressing" ϕ by consistently tuning down gravitational effects while keeping the background fixed.

We can use this new definition of $\mathcal{Q}$ to define equivalence classes of families of operators and then the notion of superficial locality[12] exactly as we do in section 7.3. All of the developments in the main text can be done in this formalism.

Topology of $\tilde{M}$

In this appendix we explain how a topology on $\tilde{M}$ can be constructed using only the boundary theory. We make no assumptions here about whether or not clumps are present. Despite appearances, the purpose of this construction is not so much to demonstrate mathematical rigor as it is to provide motivation for the statement that $\tilde{M}$, an object defined in the boundary theory, can be regarded (in the absence of clumps) as a "copy" of $\mathrm{Loc}(M)$, a region of spacetime that certainly has a nice topological structure.

The boundary theory is taken to be on a flat space which, after conformal compactification, is a cylinder. (The case where there are multiple disconnected boundaries is a straightforward generalization of the construction below.) A spatial region $R \in \mathcal{R}$ is thus bounded so its boundary, ∂R, is compact. Choose some global coordinate system on this flat spacetime (that is, fix a conformal frame), and define a Euclidean metric d between two points via geodesic (Euclidean) distance. We can now give a metric on $\mathcal{R}$ denoted by D, by defining $D(R, R')$ as the Hausdorff distance between ∂R and $\partial R'$.[13] This definition of distance is problematic in the case where $\partial R = \emptyset$. However, if $\partial R_1, \partial R_2 = \emptyset$ and $\partial R_3 \neq \emptyset$, we simply define $D(R_1, R_2) = 0$ and $D(R_1, R_3) = \infty$.

Given $\epsilon > 0$, let $B_\epsilon(R)$ be the subset of $\mathcal{R}$ consisting of regions R' with $D(R, R') < \epsilon$. A topology on $\tilde{M}$ can now be obtained by taking $P \in \tilde{M}$ and defining $U_\epsilon(P)$ as the set of points

[12] Note that the definition of superficial locality works very nicely with our new definition of $\mathcal{Q}$. If it happened, for example, that $\{\phi_\epsilon\}_{\epsilon>0}$ were a family of operators that oscillates from place to place as $\epsilon \to 0$, then we can be sure that this family would not be superficially local unless it were to oscillate within a clump.

[13] Given a metric space (S, d), the Hausdorff distance is a metric-like function that can be defined in terms of d to measure the distance between two subsets of S in a reasonable fashion. The Hausdorff distance is a legitimate metric on the collection of nonempty compact subsets of S so our definition of D provides a metric on the subset of $\mathcal{R}$ where $\partial R \neq \emptyset$ because ∂R is always compact.

$P' \in \tilde{M}$ such that for every $R \in \mathcal{Q}(P)$, there exists $R' \in \mathcal{Q}(P') \cap B_\epsilon(R)$. The collection of sets $\{U_\epsilon(P) \,\big|\, \epsilon > 0, P \in \tilde{M}\}$ forms a topological base from which a topology can be defined.

Bibliography

[1] P. A. R. Ade et al. "Detection of B-Mode Polarization at Degree Angular Scales by BICEP2". In: *Phys. Rev. Lett.* 112.24 (2014), p. 241101. DOI: 10.1103/PhysRevLett.112.241101. arXiv: 1403.3985 [astro-ph.CO].

[2] P. A. R. Ade et al. "Planck 2013 results. XXII. Constraints on inflation". In: *Astron. Astrophys.* 571 (2014), A22. DOI: 10.1051/0004-6361/201321569. arXiv: 1303.5082 [astro-ph.CO].

[3] P. A. R. Ade et al. "Planck 2015 results. XX. Constraints on inflation". In: *Astron. Astrophys.* 594 (2016), A20. DOI: 10.1051/0004-6361/201525898. arXiv: 1502.02114 [astro-ph.CO].

[4] Anthony Aguirre, Sean M. Carroll, and Matthew C. Johnson. "Out of equilibrium: understanding cosmological evolution to lower-entropy states". In: *JCAP* 1202 (2012), p. 024. DOI: 10.1088/1475-7516/2012/02/024. arXiv: 1108.0417 [hep-th].

[5] Andreas Albrecht. "Cosmic inflation and the arrow of time". In: (2002), pp. 363–401. arXiv: astro-ph/0210527 [astro-ph].

[6] Andreas Albrecht and Paul J. Steinhardt. "Cosmology for Grand Unified Theories with Radiatively Induced Symmetry Breaking". In: *Phys. Rev. Lett.* 48 (1982), pp. 1220–1223. DOI: 10.1103/PhysRevLett.48.1220.

[7] Ahmed Almheiri, Xi Dong, and Daniel Harlow. "Bulk Locality and Quantum Error Correction in AdS/CFT". In: *JHEP* 04 (2015), p. 163. DOI: 10.1007/JHEP04(2015)163. arXiv: 1411.7041 [hep-th].

[8] Ahmed Almheiri, Xi Dong, and Brian Swingle. "Linearity of Holographic Entanglement Entropy". In: *JHEP* 02 (2017), p. 074. DOI: 10.1007/JHEP02(2017)074. arXiv: 1606.04537 [hep-th].

[9] Ahmed Almheiri et al. "An Apologia for Firewalls". In: *JHEP* 09 (2013), p. 018. DOI: 10.1007/JHEP09(2013)018. arXiv: 1304.6483 [hep-th].

[10] Ahmed Almheiri et al. "Black Holes: Complementarity or Firewalls?" In: *JHEP* 02 (2013), p. 062. DOI: 10.1007/JHEP02(2013)062. arXiv: 1207.3123 [hep-th].

[11] Asimina Arvanitaki et al. "String Axiverse". In: *Phys. Rev.* D81 (2010), p. 123530. DOI: 10.1103/PhysRevD.81.123530. arXiv: 0905.4720 [hep-th].

[12] Abhay Ashtekar and Gregory J. Galloway. "Some uniqueness results for dynamical horizons". In: *Adv. Theor. Math. Phys.* 9.1 (2005), pp. 1–30. DOI: 10.4310/ATMP.2005.v9.n1.a1. arXiv: gr-qc/0503109 [gr-qc].

[13] Abhay Ashtekar and Badri Krishnan. "Dynamical horizons and their properties". In: *Phys. Rev.* D68 (2003), p. 104030. DOI: 10.1103/PhysRevD.68.104030. arXiv: gr-qc/0308033 [gr-qc].

[14] Abhay Ashtekar and Badri Krishnan. "Dynamical horizons: Energy, angular momentum, fluxes and balance laws". In: *Phys. Rev. Lett.* 89 (2002), p. 261101. DOI: 10.1103/PhysRevLett.89.261101. arXiv: gr-qc/0207080 [gr-qc].

[15] Vijay Balasubramanian et al. "Bulk curves from boundary data in holography". In: *Phys. Rev.* D89.8 (2014), p. 086004. DOI: 10.1103/PhysRevD.89.086004. arXiv: 1310.4204 [hep-th].

[16] Vijay Balasubramanian et al. "Entwinement and the emergence of spacetime". In: *JHEP* 01 (2015), p. 048. DOI: 10.1007/JHEP01(2015)048. arXiv: 1406.5859 [hep-th].

[17] Vijay Balasubramanian et al. "The entropy of a hole in spacetime". In: *JHEP* 10 (2013), p. 220. DOI: 10.1007/JHEP10(2013)220. arXiv: 1305.0856 [hep-th].

[18] Tom Banks et al. "AdS dynamics from conformal field theory". In: (1998). arXiv: hep-th/9808016 [hep-th].

[19] James M. Bardeen. "Black Holes Do Evaporate Thermally". In: *Phys. Rev. Lett.* 46 (1981), pp. 382–385. DOI: 10.1103/PhysRevLett.46.382.

[20] James M. Bardeen, B. Carter, and S. W. Hawking. "The Four laws of black hole mechanics". In: *Commun. Math. Phys.* 31 (1973), pp. 161–170. DOI: 10.1007/BF01645742.

[21] J. D. Bekenstein. "Black holes and the second law". In: *Lett. Nuovo Cim.* 4 (1972), pp. 737–740. DOI: 10.1007/BF02757029.

[22] Jacob D. Bekenstein. "Black holes and entropy". In: *Phys. Rev.* D7 (1973), pp. 2333–2346. DOI: 10.1103/PhysRevD.7.2333.

[23] Maria Beltran, Juan Garcia-Bellido, and Julien Lesgourgues. "Isocurvature bounds on axions revisited". In: *Phys. Rev.* D75 (2007), p. 103507. DOI: 10.1103/PhysRevD.75.103507. arXiv: hep-ph/0606107 [hep-ph].

[24] C. Beny, A. Kempf, and D. W. Kribs. "Generalization of Quantum Error Correction via the Heisenberg Picture". In: *Physical Review Letters* 98.10, 100502 (Mar. 2007), p. 100502. DOI: 10.1103/PhysRevLett.98.100502. eprint: quant-ph/0608071.

[25] C. Beny, A. Kempf, and D. W. Kribs. "Quantum error correction of observables". In: *Physical Review A* 76.4, 042303 (Oct. 2007), p. 042303. DOI: 10.1103/PhysRevA.76.042303. arXiv: 0705.1574 [quant-ph].

[26] Luca Bombelli et al. "A Quantum Source of Entropy for Black Holes". In: *Phys. Rev.* D34 (1986), pp. 373–383. DOI: 10.1103/PhysRevD.34.373.

[27] Raphael Bousso. "A Covariant entropy conjecture". In: *JHEP* 07 (1999), p. 004. DOI: 10.1088/1126-6708/1999/07/004. arXiv: hep-th/9905177 [hep-th].

[28] Raphael Bousso. "Holography in general space-times". In: *JHEP* 06 (1999), p. 028. DOI: 10.1088/1126-6708/1999/06/028. arXiv: hep-th/9906022 [hep-th].

[29] Raphael Bousso. "The Holographic principle". In: *Rev. Mod. Phys.* 74 (2002), pp. 825–874. DOI: 10.1103/RevModPhys.74.825. arXiv: hep-th/0203101 [hep-th].

[30] Raphael Bousso and Netta Engelhardt. "Generalized Second Law for Cosmology". In: *Phys. Rev.* D93.2 (2016), p. 024025. DOI: 10.1103/PhysRevD.93.024025. arXiv: 1510.02099 [hep-th].

[31] Raphael Bousso and Netta Engelhardt. "New Area Law in General Relativity". In: *Phys. Rev. Lett.* 115.8 (2015), p. 081301. DOI: 10.1103/PhysRevLett.115.081301. arXiv: 1504.07627 [hep-th].

[32] Raphael Bousso and Netta Engelhardt. "Proof of a New Area Law in General Relativity". In: *Phys. Rev.* D92.4 (2015), p. 044031. DOI: 10.1103/PhysRevD.92.044031. arXiv: 1504.07660 [gr-qc].

[33] Raphael Bousso, Ben Freivogel, and Stefan Leichenauer. "Saturating the holographic entropy bound". In: *Phys. Rev.* D82 (2010), p. 084024. DOI: 10.1103/PhysRevD.82.084024. arXiv: 1003.3012 [hep-th].

[34] Raphael Bousso and Joseph Polchinski. "Quantization of four form fluxes and dynamical neutralization of the cosmological constant". In: *JHEP* 06 (2000), p. 006. DOI: 10.1088/1126-6708/2000/06/006. arXiv: hep-th/0004134 [hep-th].

[35] Raphael Bousso and Leonard Susskind. "The Multiverse Interpretation of Quantum Mechanics". In: *Phys. Rev.* D85 (2012), p. 045007. DOI: 10.1103/PhysRevD.85.045007. arXiv: 1105.3796 [hep-th].

[36] Raphael Bousso et al. "Null Geodesics, Local CFT Operators and AdS/CFT for Subregions". In: *Phys. Rev.* D88 (2013), p. 064057. DOI: 10.1103/PhysRevD.88.064057. arXiv: 1209.4641 [hep-th].

[37] Robert H. Brandenberger. "On the Decay of Cosmic String Loops". In: *Nucl. Phys.* B293 (1987), pp. 812–828. DOI: 10.1016/0550-3213(87)90092-7.

[38] Samuel L. Braunstein, Stefano Pirandola, and Karol Zyczkowski. "Better Late than Never: Information Retrieval from Black Holes". In: *Phys. Rev. Lett.* 110.10 (2013), p. 101301. DOI: 10.1103/PhysRevLett.110.101301. arXiv: 0907.1190 [quant-ph].

[39] Adam R. Brown. "Tensile Strength and the Mining of Black Holes". In: *Phys. Rev. Lett.* 111.21 (2013), p. 211301. DOI: 10.1103/PhysRevLett.111.211301. arXiv: 1207.3342 [gr-qc].

[40] Dmitry Budker et al. "Proposal for a Cosmic Axion Spin Precession Experiment (CASPEr)". In: *Phys. Rev.* X4.2 (2014), p. 021030. DOI: 10.1103/PhysRevX.4.021030. arXiv: 1306.6089 [hep-ph].

[41] Sean M. Carroll and Jennifer Chen. "Spontaneous inflation and the origin of the arrow of time". In: (2004). arXiv: hep-th/0410270 [hep-th].

[42] J. A. Casas and C. Munoz. "A Natural solution to the mu problem". In: *Phys. Lett.* B306 (1993), pp. 288–294. DOI: 10.1016/0370-2693(93)90081-R. arXiv: hep-ph/9302227 [hep-ph].

[43] Kiwoon Choi et al. "Diluting the inflationary axion fluctuation by a stronger QCD in the early Universe". In: *Phys. Lett.* B750 (2015), pp. 26–30. DOI: 10.1016/j.physletb.2015.08.041. arXiv: 1505.00306 [hep-ph].

[44] Sidney R. Coleman and Frank De Luccia. "Gravitational Effects on and of Vacuum Decay". In: *Phys. Rev.* D21 (1980), p. 3305. DOI: 10.1103/PhysRevD.21.3305.

[45] Bartlomiej Czech, Xi Dong, and James Sully. "Holographic Reconstruction of General Bulk Surfaces". In: *JHEP* 11 (2014), p. 015. DOI: 10.1007/JHEP11(2014)015. arXiv: 1406.4889 [hep-th].

[46] Bartlomiej Czech et al. "Integral Geometry and Holography". In: *JHEP* 10 (2015), p. 175. DOI: 10.1007/JHEP10(2015)175. arXiv: 1505.05515 [hep-th].

[47] Bart?omiej Czech and Lampros Lamprou. "Holographic definition of points and distances". In: *Phys. Rev.* D90 (2014), p. 106005. DOI: 10.1103/PhysRevD.90.106005. arXiv: 1409.4473 [hep-th].

[48] Bryce S. DeWitt. "Quantum Theory of Gravity. 1. The Canonical Theory". In: *Phys. Rev.* 160 (1967), pp. 1113–1148. DOI: 10.1103/PhysRev.160.1113.

[49] Xi Dong, Daniel Harlow, and Aron C. Wall. "Reconstruction of Bulk Operators within the Entanglement Wedge in Gauge-Gravity Duality". In: *Phys. Rev. Lett.* 117.2 (2016), p. 021601. DOI: 10.1103/PhysRevLett.117.021601. arXiv: 1601.05416 [hep-th].

[50] Xi Dong, Aitor Lewkowycz, and Mukund Rangamani. "Deriving covariant holographic entanglement". In: *JHEP* 11 (2016), p. 028. DOI: 10.1007/JHEP11(2016)028. arXiv: 1607.07506 [hep-th].

[51] William Donnelly and Steven B. Giddings. "Diffeomorphism-invariant observables and their nonlocal algebra". In: *Phys. Rev.* D93.2 (2016). [Erratum: Phys. Rev.D94,no.2,029903(2016 p. 024030. DOI: 10.1103/PhysRevD.94.029903,10.1103/PhysRevD.93.024030. arXiv: 1507.07921 [hep-th].

[52] William Donnelly and Steven B. Giddings. "Observables, gravitational dressing, and obstructions to locality and subsystems". In: *Phys. Rev.* D94.10 (2016), p. 104038. DOI: 10.1103/PhysRevD.94.104038. arXiv: 1607.01025 [hep-th].

[53] Michael R. Douglas. "The Statistics of string / M theory vacua". In: *JHEP* 05 (2003), p. 046. DOI: 10.1088/1126-6708/2003/05/046. arXiv: hep-th/0303194 [hep-th].

[54] G. R. Dvali. "Removing the cosmological bound on the axion scale". In: (1995). arXiv: hep-ph/9505253 [hep-ph].

[55] Gia Dvali. "Black Holes and Large N Species Solution to the Hierarchy Problem". In: *Fortsch. Phys.* 58 (2010), pp. 528–536. DOI: 10.1002/prop.201000009. arXiv: 0706.2050 [hep-th].

[56] Lisa Dyson, Matthew Kleban, and Leonard Susskind. "Disturbing implications of a cosmological constant". In: *JHEP* 10 (2002), p. 011. DOI: 10.1088/1126-6708/2002/10/011. arXiv: hep-th/0208013 [hep-th].

[57] Netta Engelhardt. "Into the Bulk: A Covariant Approach". In: *Phys. Rev.* D95.6 (2017), p. 066005. DOI: 10.1103/PhysRevD.95.066005. arXiv: 1610.08516 [hep-th].

[58] Netta Engelhardt and Sebastian Fischetti. "Causal Density Matrices". In: *Phys. Rev.* D95.12 (2017), p. 126012. DOI: 10.1103/PhysRevD.95.126012. arXiv: 1703.05328 [hep-th].

[59] Netta Engelhardt and Gary T. Horowitz. "New Insights into Quantum Gravity from Gauge/gravity Duality". In: *Int. J. Mod. Phys.* D25.12 (2016), p. 1643002. DOI: 10.1142/S0218271816430021. arXiv: 1605.04335 [hep-th].

[60] Netta Engelhardt and Gary T. Horowitz. "Recovering the spacetime metric from a holographic dual". In: *Adv. Theor. Math. Phys.* 21 (2017), pp. 1635–1653. DOI: 10.4310/ATMP.2017.v21.n7.a2. arXiv: 1612.00391 [hep-th].

[61] Netta Engelhardt and Gary T. Horowitz. "Towards a Reconstruction of General Bulk Metrics". In: *Class. Quant. Grav.* 34.1 (2017), p. 015004. DOI: 10.1088/1361-6382/34/1/015004. arXiv: 1605.01070 [hep-th].

[62] Netta Engelhardt and Aron C. Wall. "Extremal Surface Barriers". In: *JHEP* 03 (2014), p. 068. DOI: 10.1007/JHEP03(2014)068. arXiv: 1312.3699 [hep-th].

[63] Netta Engelhardt and Aron C. Wall. "Quantum Extremal Surfaces: Holographic Entanglement Entropy beyond the Classical Regime". In: *JHEP* 01 (2015), p. 073. DOI: 10.1007/JHEP01(2015)073. arXiv: 1408.3203 [hep-th].

[64] Thomas Faulkner, Aitor Lewkowycz, and Juan Maldacena. "Quantum corrections to holographic entanglement entropy". In: *JHEP* 11 (2013), p. 074. DOI: 10.1007/JHEP11(2013)074. arXiv: 1307.2892 [hep-th].

[65] W. Fischler and Leonard Susskind. "Holography and cosmology". In: (1998). arXiv: hep-th/9806039 [hep-th].

[66] Willy Fischler and John Preskill. "DYON - AXION DYNAMICS". In: *Phys. Lett.* 125B (1983), pp. 165–170. DOI: 10.1016/0370-2693(83)91260-1.

[67] Patrick Fox, Aaron Pierce, and Scott D. Thomas. "Probing a QCD string axion with precision cosmological measurements". In: (2004). arXiv: hep-th/0409059 [hep-th].

[68] Michael Freedman and Matthew Headrick. "Bit threads and holographic entanglement". In: *Commun. Math. Phys.* 352.1 (2017), pp. 407–438. DOI: 10.1007/s00220-016-2796-3. arXiv: 1604.00354 [hep-th].

[69] Mirah Gary, Steven B. Giddings, and Joao Penedones. "Local bulk S-matrix elements and CFT singularities". In: *Phys. Rev.* D80 (2009), p. 085005. DOI: 10.1103/PhysRevD.80.085005. arXiv: 0903.4437 [hep-th].

[70] G. W. Gibbons and S. W. Hawking. "Cosmological Event Horizons, Thermodynamics, and Particle Creation". In: *Phys. Rev.* D15 (1977), pp. 2738–2751. DOI: 10.1103/PhysRevD.15.2738.

[71] G. F. Giudice and A. Masiero. "A Natural Solution to the mu Problem in Supergravity Theories". In: *Phys. Lett.* B206 (1988), pp. 480–484. DOI: 10.1016/0370-2693(88)91613-9.

[72] Peter W. Graham and Surjeet Rajendran. "Axion Dark Matter Detection with Cold Molecules". In: *Phys. Rev.* D84 (2011), p. 055013. DOI: 10.1103/PhysRevD.84.055013. arXiv: 1101.2691 [hep-ph].

[73] Peter W. Graham and Surjeet Rajendran. "New Observables for Direct Detection of Axion Dark Matter". In: *Phys. Rev.* D88 (2013), p. 035023. DOI: 10.1103/PhysRevD.88.035023. arXiv: 1306.6088 [hep-ph].

[74] S. S. Gubser, Igor R. Klebanov, and Alexander M. Polyakov. "Gauge theory correlators from noncritical string theory". In: *Phys. Lett.* B428 (1998), pp. 105–114. DOI: 10.1016/S0370-2693(98)00377-3. arXiv: hep-th/9802109 [hep-th].

[75] Alan H. Guth. "The Inflationary Universe: A Possible Solution to the Horizon and Flatness Problems". In: *Phys. Rev.* D23 (1981), pp. 347–356. DOI: 10.1103/PhysRevD.23.347.

[76] Alan H. Guth, David I. Kaiser, and Yasunori Nomura. "Inflationary paradigm after Planck 2013". In: *Phys. Lett.* B733 (2014), pp. 112–119. DOI: 10.1016/j.physletb.2014.03.020. arXiv: 1312.7619 [astro-ph.CO].

[77] Alan H. Guth and S. Y. Pi. "Fluctuations in the New Inflationary Universe". In: *Phys. Rev. Lett.* 49 (1982), pp. 1110–1113. DOI: 10.1103/PhysRevLett.49.1110.

[78] Alan H. Guth and Erick J. Weinberg. "Could the Universe Have Recovered from a Slow First Order Phase Transition?" In: *Nucl. Phys.* B212 (1983), pp. 321–364. DOI: 10.1016/0550-3213(83)90307-3.

[79] Hal M. Haggard and Carlo Rovelli. "Quantum-gravity effects outside the horizon spark black to white hole tunneling". In: *Phys. Rev.* D92.10 (2015), p. 104020. DOI: 10.1103/PhysRevD.92.104020. arXiv: 1407.0989 [gr-qc].

[80] Alex Hamilton et al. "Holographic representation of local bulk operators". In: *Phys. Rev.* D74 (2006), p. 066009. DOI: 10.1103/PhysRevD.74.066009. arXiv: hep-th/0606141 [hep-th].

[81] Daniel Harlow. "The Ryu-Takayanagi Formula from Quantum Error Correction". In: *Commun. Math. Phys.* 354.3 (2017), pp. 865–912. DOI: 10.1007/s00220-017-2904-z. arXiv: 1607.03901 [hep-th].

[82] J. B. Hartle and S. W. Hawking. "Path Integral Derivation of Black Hole Radiance". In: *Phys. Rev.* D13 (1976), pp. 2188–2203. DOI: 10.1103/PhysRevD.13.2188.

[83] Thomas Hartman and Juan Maldacena. "Time Evolution of Entanglement Entropy from Black Hole Interiors". In: *JHEP* 05 (2013), p. 014. DOI: 10.1007/JHEP05(2013)014. arXiv: 1303.1080 [hep-th].

[84] S. W. Hawking. "Black hole explosions". In: *Nature* 248 (1974), pp. 30–31. DOI: 10.1038/248030a0.

[85] S. W. Hawking. "Breakdown of Predictability in Gravitational Collapse". In: *Phys. Rev.* D14 (1976), pp. 2460–2473. DOI: 10.1103/PhysRevD.14.2460.

[86] S. W. Hawking. "Particle Creation by Black Holes". In: *Commun. Math. Phys.* 43 (1975). [,167(1975)], pp. 199–220. DOI: 10.1007/BF02345020,10.1007/BF01608497.

[87] S. W. Hawking. "The Development of Irregularities in a Single Bubble Inflationary Universe". In: *Phys. Lett.* 115B (1982), p. 295. DOI: 10.1016/0370-2693(82)90373-2.

[88] Patrick Hayden and John Preskill. "Black holes as mirrors: Quantum information in random subsystems". In: *JHEP* 09 (2007), p. 120. DOI: 10.1088/1126-6708/2007/09/120. arXiv: 0708.4025 [hep-th].

[89] S. A. Hayward. "General laws of black hole dynamics". In: *Phys. Rev.* D49 (1994), pp. 6467–6474. DOI: 10.1103/PhysRevD.49.6467.

[90] Sean A. Hayward. "Unified first law of black hole dynamics and relativistic thermodynamics". In: *Class. Quant. Grav.* 15 (1998), pp. 3147–3162. DOI: 10.1088/0264-9381/15/10/017. arXiv: gr-qc/9710089 [gr-qc].

[91] Matthew Headrick, Robert C. Myers, and Jason Wien. "Holographic Holes and Differential Entropy". In: *JHEP* 10 (2014), p. 149. DOI: 10.1007/JHEP10(2014)149. arXiv: 1408.4770 [hep-th].

[92] Matthew Headrick and Tadashi Takayanagi. "A Holographic proof of the strong subadditivity of entanglement entropy". In: *Phys. Rev.* D76 (2007), p. 106013. DOI: 10.1103/PhysRevD.76.106013. arXiv: 0704.3719 [hep-th].

[93] Matthew Headrick et al. "Causality & holographic entanglement entropy". In: *JHEP* 12 (2014), p. 162. DOI: 10.1007/JHEP12(2014)162. arXiv: 1408.6300 [hep-th].

[94] Idse Heemskerk et al. "Bulk and Transhorizon Measurements in AdS/CFT". In: *JHEP* 10 (2012), p. 165. DOI: 10.1007/JHEP10(2012)165. arXiv: 1201.3664 [hep-th].

[95] Idse Heemskerk et al. "Holography from Conformal Field Theory". In: *JHEP* 10 (2009), p. 079. DOI: 10.1088/1126-6708/2009/10/079. arXiv: 0907.0151 [hep-th].

[96] Mark P Hertzberg, Max Tegmark, and Frank Wilczek. "Axion Cosmology and the Energy Scale of Inflation". In: *Phys. Rev.* D78 (2008), p. 083507. DOI: 10.1103/PhysRevD.78.083507. arXiv: 0807.1726 [astro-ph].

[97] Veronika E. Hubeny and Henry Maxfield. "Holographic probes of collapsing black holes". In: *JHEP* 03 (2014), p. 097. DOI: 10.1007/JHEP03(2014)097. arXiv: 1312.6887 [hep-th].

[98] Veronika E. Hubeny and Mukund Rangamani. "Causal Holographic Information". In: *JHEP* 06 (2012), p. 114. DOI: 10.1007/JHEP06(2012)114. arXiv: 1204.1698 [hep-th].

[99] Veronika E. Hubeny, Mukund Rangamani, and Tadashi Takayanagi. "A Covariant holographic entanglement entropy proposal". In: *JHEP* 07 (2007), p. 062. DOI: 10.1088/1126-6708/2007/07/062. arXiv: 0705.0016 [hep-th].

[100] Ted Jacobson. "Thermodynamics of space-time: The Einstein equation of state". In: *Phys. Rev. Lett.* 75 (1995), pp. 1260–1263. DOI: 10.1103/PhysRevLett.75.1260. arXiv: gr-qc/9504004 [gr-qc].

[101] Daniel L. Jafferis and S. Josephine Suh. "The Gravity Duals of Modular Hamiltonians". In: *JHEP* 09 (2016), p. 068. DOI: 10.1007/JHEP09(2016)068. arXiv: 1412.8465 [hep-th].

[102] Daniel L. Jafferis et al. "Relative entropy equals bulk relative entropy". In: *JHEP* 06 (2016), p. 004. DOI: 10.1007/JHEP06(2016)004. arXiv: 1512.06431 [hep-th].

[103] Daniel Kabat, Gilad Lifschytz, and David A. Lowe. "Constructing local bulk observables in interacting AdS/CFT". In: *Phys. Rev.* D83 (2011), p. 106009. DOI: 10.1103/PhysRevD.83.106009. arXiv: 1102.2910 [hep-th].

[104] Shamit Kachru et al. "De Sitter vacua in string theory". In: *Phys. Rev.* D68 (2003), p. 046005. DOI: 10.1103/PhysRevD.68.046005. arXiv: hep-th/0301240 [hep-th].

[105] Marc Kamionkowski, Arthur Kosowsky, and Albert Stebbins. "A Probe of primordial gravity waves and vorticity". In: *Phys. Rev. Lett.* 78 (1997), pp. 2058–2061. DOI: 10.1103/PhysRevLett.78.2058. arXiv: astro-ph/9609132 [astro-ph].

[106] Marc Kamionkowski and John March-Russell. "Planck scale physics and the Peccei-Quinn mechanism". In: *Phys. Lett.* B282 (1992), pp. 137–141. DOI: 10.1016/0370-2693(92)90492-M. arXiv: hep-th/9202003 [hep-th].

[107] Junhai Kang, Markus A. Luty, and Salah Nasri. "The Relic abundance of long-lived heavy colored particles". In: *JHEP* 09 (2008), p. 086. DOI: 10.1088/1126-6708/2008/09/086. arXiv: hep-ph/0611322 [hep-ph].

[108] David B. Kaplan and Kathryn M. Zurek. "Exotic axions". In: *Phys. Rev. Lett.* 96 (2006), p. 041301. DOI: 10.1103/PhysRevLett.96.041301. arXiv: hep-ph/0507236 [hep-ph].

[109] Masahiro Kawasaki, Naoya Kitajima, and Fuminobu Takahashi. "Relaxing Isocurvature Bounds on String Axion Dark Matter". In: *Phys. Lett.* B737 (2014), pp. 178–184. DOI: 10.1016/j.physletb.2014.08.017. arXiv: 1406.0660 [hep-ph].

[110] T. W. B. Kibble. "Topology of Cosmic Domains and Strings". In: *J. Phys.* A9 (1976), pp. 1387–1398. DOI: 10.1088/0305-4470/9/8/029.

[111] Naoya Kitajima and Fuminobu Takahashi. "Resonant conversions of QCD axions into hidden axions and suppressed isocurvature perturbations". In: *JCAP* 1501.01 (2015), p. 032. DOI: 10.1088/1475-7516/2015/01/032. arXiv: 1411.2011 [hep-ph].

[112] A. Kogut et al. "The Primordial Inflation Explorer (PIXIE): A Nulling Polarimeter for Cosmic Microwave Background Observations". In: *JCAP* 1107 (2011), p. 025. DOI: 10.1088/1475-7516/2011/07/025. arXiv: 1105.2044 [astro-ph.CO].

[113] Aitor Lewkowycz and Juan Maldacena. "Generalized gravitational entropy". In: *JHEP* 08 (2013), p. 090. DOI: 10.1007/JHEP08(2013)090. arXiv: 1304.4926 [hep-th].

[114] Wei Li and Tadashi Takayanagi. "Holography and Entanglement in Flat Spacetime". In: *Phys. Rev. Lett.* 106 (2011), p. 141301. DOI: 10.1103/PhysRevLett.106.141301. arXiv: 1010.3700 [hep-th].

[115] Andrei D. Linde. "A New Inflationary Universe Scenario: A Possible Solution of the Horizon, Flatness, Homogeneity, Isotropy and Primordial Monopole Problems". In: *Phys. Lett.* 108B (1982), pp. 389–393. DOI: 10.1016/0370-2693(82)91219-9.

[116] Andrei D. Linde. "ETERNAL CHAOTIC INFLATION". In: *Mod. Phys. Lett.* A1 (1986), p. 81. DOI: 10.1142/S0217732386000129.

[117] Andrei D. Linde. "Eternally Existing Selfreproducing Chaotic Inflationary Universe". In: *Phys. Lett.* B175 (1986), pp. 395–400. DOI: 10.1016/0370-2693(86)90611-8.

[118] Andrei D. Linde. "GENERATION OF ISOTHERMAL DENSITY PERTURBATIONS IN THE INFLATIONARY UNIVERSE". In: *JETP Lett.* 40 (1984). [Pisma Zh. Eksp. Teor. Fiz.40,496(1984)], pp. 1333–1336.

[119] Andrei D. Linde. "Generation of Isothermal Density Perturbations in the Inflationary Universe". In: *Phys. Lett.* 158B (1985), pp. 375–380. DOI: 10.1016/0370-2693(85)90436-8.

[120] Andrei D. Linde. "Inflation and Axion Cosmology". In: *Phys. Lett.* B201 (1988), pp. 437–439. DOI: 10.1016/0370-2693(88)90597-7.

[121] Katherine J. Mack. "Axions, Inflation and the Anthropic Principle". In: *JCAP* 1107 (2011), p. 021. DOI: 10.1088/1475-7516/2011/07/021. arXiv: 0911.0421 [astro-ph.CO].

[122] Juan Martin Maldacena. "The Large N limit of superconformal field theories and supergravity". In: *Int. J. Theor. Phys.* 38 (1999). [Adv. Theor. Math. Phys.2,231(1998)], pp. 1113–1133. DOI: 10.1023/A:1026654312961,10.4310/ATMP.1998.v2.n2.a1. arXiv: hep-th/9711200 [hep-th].

[123] Juan Maldacena, David Simmons-Duffin, and Alexander Zhiboedov. "Looking for a bulk point". In: *JHEP* 01 (2017), p. 013. DOI: 10.1007/JHEP01(2017)013. arXiv: 1509.03612 [hep-th].

[124] Juan Maldacena and Leonard Susskind. "Cool horizons for entangled black holes". In: *Fortsch. Phys.* 61 (2013), pp. 781–811. DOI: 10.1002/prop.201300020. arXiv: 1306.0533 [hep-th].

[125] Donald Marolf. "Black Holes, AdS, and CFTs". In: *Gen. Rel. Grav.* 41 (2009), pp. 903–917. DOI: 10.1007/s10714-008-0749-7. arXiv: 0810.4886 [gr-qc].

[126] Donald Marolf and Joseph Polchinski. "Gauge/Gravity Duality and the Black Hole Interior". In: *Phys. Rev. Lett.* 111 (2013), p. 171301. DOI: 10.1103/PhysRevLett.111.171301. arXiv: 1307.4706 [hep-th].

[127] Samir D. Mathur. "The Information paradox: A Pedagogical introduction". In: *Class. Quant. Grav.* 26 (2009), p. 224001. DOI: 10.1088/0264-9381/26/22/224001. arXiv: 0909.1038 [hep-th].

[128] Ian A. Morrison. "Boundary-to-bulk maps for AdS causal wedges and the Reeh-Schlieder property in holography". In: *JHEP* 05 (2014), p. 053. DOI: 10.1007/JHEP05(2014)053. arXiv: 1403.3426 [hep-th].

[129] Hitoshi Murayama and Jing Shu. "Topological Dark Matter". In: *Phys. Lett.* B686 (2010), pp. 162–165. DOI: 10.1016/j.physletb.2010.02.037. arXiv: 0905.1720 [hep-ph].

[130] E. T. Newman. "Heaven and Its Properties". In: *Gen. Rel. Grav.* 7 (1976), pp. 107–111. DOI: 10.1007/BF00762018.

[131] Yasunori Nomura. "A Note on Boltzmann Brains". In: *Phys. Lett.* B749 (2015), pp. 514–518. DOI: 10.1016/j.physletb.2015.08.029. arXiv: 1502.05401 [hep-th].

[132] Yasunori Nomura. "Physical Theories, Eternal Inflation, and Quantum Universe". In: *JHEP* 11 (2011), p. 063. DOI: 10.1007/JHEP11(2011)063. arXiv: 1104.2324 [hep-th].

[133] Yasunori Nomura. "Quantum Mechanics, Spacetime Locality, and Gravity". In: *Found. Phys.* 43 (2013), pp. 978–1007. DOI: 10.1007/s10701-013-9729-1. arXiv: 1110.4630 [hep-th].

[134] Yasunori Nomura. "The Static Quantum Multiverse". In: *Phys. Rev.* D86 (2012), p. 083505. DOI: 10.1103/PhysRevD.86.083505. arXiv: 1205.5550 [hep-th].

[135] Yasunori Nomura and Nico Salzetta. "Why Firewalls Need Not Exist". In: *Phys. Lett.* B761 (2016), pp. 62–69. DOI: 10.1016/j.physletb.2016.08.003. arXiv: 1602.07673 [hep-th].

[136] Yasunori Nomura, Fabio Sanches, and Sean J. Weinberg. "Black Hole Interior in Quantum Gravity". In: *Phys. Rev. Lett.* 114 (2015), p. 201301. DOI: 10.1103/PhysRevLett.114.201301. arXiv: 1412.7539 [hep-th].

[137] Yasunori Nomura, Fabio Sanches, and Sean J. Weinberg. "Relativeness in Quantum Gravity: Limitations and Frame Dependence of Semiclassical Descriptions". In: *JHEP* 04 (2015), p. 158. DOI: 10.1007/JHEP04(2015)158. arXiv: 1412.7538 [hep-th].

[138] Yasunori Nomura and Jaime Varela. "A Note on (No) Firewalls: The Entropy Argument". In: *JHEP* 07 (2013), p. 124. DOI: 10.1007/JHEP07(2013)124. arXiv: 1211.7033 [hep-th].

[139] Yasunori Nomura, Jaime Varela, and Sean J. Weinberg. "Black Holes, Information, and Hilbert Space for Quantum Gravity". In: *Phys. Rev.* D87 (2013), p. 084050. DOI: 10.1103/PhysRevD.87.084050. arXiv: 1210.6348 [hep-th].

[140] Yasunori Nomura, Jaime Varela, and Sean J. Weinberg. "Low Energy Description of Quantum Gravity and Complementarity". In: *Phys. Lett.* B733 (2014), pp. 126–133. DOI: 10.1016/j.physletb.2014.04.027. arXiv: 1304.0448 [hep-th].

[141] Yasunori Nomura and Sean J. Weinberg. "Black Holes, Entropies, and Semiclassical Spacetime in Quantum Gravity". In: *JHEP* 10 (2014), p. 185. DOI: 10.1007/JHEP10(2014)185. arXiv: 1406.1505 [hep-th].

[142] Yasunori Nomura and Sean J. Weinberg. "Entropy of a vacuum: What does the covariant entropy count?" In: *Phys. Rev.* D90.10 (2014), p. 104003. DOI: 10.1103/PhysRevD.90.104003. arXiv: 1310.7564 [hep-th].

[143] Yasunori Nomura et al. "Spacetime Equals Entanglement". In: *Phys. Lett.* B763 (2016), pp. 370–374. DOI: 10.1016/j.physletb.2016.10.045. arXiv: 1607.02508 [hep-th].

[144] Don N. Page. "Average entropy of a subsystem". In: *Phys. Rev. Lett.* 71 (1993), pp. 1291–1294. DOI: 10.1103/PhysRevLett.71.1291. arXiv: gr-qc/9305007 [gr-qc].

[145] Don N. Page. "Information in black hole radiation". In: *Phys. Rev. Lett.* 71 (1993), pp. 3743–3746. DOI: 10.1103/PhysRevLett.71.3743. arXiv: hep-th/9306083 [hep-th].

[146] Don N. Page. "IS BLACK HOLE EVAPORATION PREDICTABLE?" In: *Phys. Rev. Lett.* 44 (1980), p. 301. DOI: 10.1103/PhysRevLett.44.301.

[147] Don N. Page. "Is our universe likely to decay within 20 billion years?" In: *Phys. Rev.* D78 (2008), p. 063535. DOI: 10.1103/PhysRevD.78.063535. arXiv: hep-th/0610079 [hep-th].

[148] Don N. Page. "Particle Emission Rates from a Black Hole: Massless Particles from an Uncharged, Nonrotating Hole". In: *Phys. Rev.* D13 (1976), pp. 198–206. DOI: 10.1103/PhysRevD.13.198.

[149] Don N. Page and William K. Wootters. "EVOLUTION WITHOUT EVOLUTION: DYNAMICS DESCRIBED BY STATIONARY OBSERVABLES". In: *Phys. Rev.* D27 (1983), p. 2885. DOI: 10.1103/PhysRevD.27.2885.

[150] Kyriakos Papadodimas and Suvrat Raju. "An Infalling Observer in AdS/CFT". In: *JHEP* 10 (2013), p. 212. DOI: 10.1007/JHEP10(2013)212. arXiv: 1211.6767 [hep-th].

[151] Kyriakos Papadodimas and Suvrat Raju. "Remarks on the necessity and implications of state-dependence in the black hole interior". In: *Phys. Rev.* D93.8 (2016), p. 084049. DOI: 10.1103/PhysRevD.93.084049. arXiv: 1503.08825 [hep-th].

[152] Alejandro Perez. "No firewalls in quantum gravity: the role of discreteness of quantum geometry in resolving the information loss paradox". In: *Class. Quant. Grav.* 32.8 (2015), p. 084001. DOI: 10.1088/0264-9381/32/8/084001. arXiv: 1410.7062 [gr-qc].

[153] John Preskill. "Cosmological Production of Superheavy Magnetic Monopoles". In: *Phys. Rev. Lett.* 43 (1979), p. 1365. DOI: 10.1103/PhysRevLett.43.1365.

[154] John Preskill. "Do black holes destroy information?" In: *International Symposium on Black holes, Membranes, Wormholes and Superstrings Woodlands, Texas, January 16-18, 1992*. 1992, pp. 22–39. arXiv: hep-th/9209058 [hep-th].

[155] Carlo Rovelli. "Black hole entropy from loop quantum gravity". In: *Phys. Rev. Lett.* 77 (1996), pp. 3288–3291. DOI: 10.1103/PhysRevLett.77.3288. arXiv: gr-qc/9603063 [gr-qc].

[156] Shinsei Ryu and Tadashi Takayanagi. "Aspects of Holographic Entanglement Entropy". In: *JHEP* 08 (2006), p. 045. DOI: 10.1088/1126-6708/2006/08/045. arXiv: hep-th/0605073 [hep-th].

[157] Shinsei Ryu and Tadashi Takayanagi. "Holographic derivation of entanglement entropy from AdS/CFT". In: *Phys. Rev. Lett.* 96 (2006), p. 181602. DOI: 10.1103/PhysRevLett.96.181602. arXiv: hep-th/0603001 [hep-th].

[158] Fabio Sanches and Sean J. Weinberg. "Holographic entanglement entropy conjecture for general spacetimes". In: *Phys. Rev.* D94.8 (2016), p. 084034. DOI: 10.1103/PhysRevD.94.084034. arXiv: 1603.05250 [hep-th].

[159] Fabio Sanches and Sean J. Weinberg. "Refinement of the Bousso-Engelhardt Area Law". In: *Phys. Rev.* D94.2 (2016), p. 021502. DOI: 10.1103/PhysRevD.94.021502. arXiv: 1604.04919 [hep-th].

[160] D. Seckel and Michael S. Turner. "Isothermal Density Perturbations in an Axion Dominated Inflationary Universe". In: *Phys. Rev.* D32 (1985), p. 3178. DOI: 10.1103/PhysRevD.32.3178.

[161] Yasuhiro Sekino and Leonard Susskind. "Fast Scramblers". In: *JHEP* 10 (2008), p. 065. DOI: 10.1088/1126-6708/2008/10/065. arXiv: 0808.2096 [hep-th].

[162] Uros Seljak, Anze Slosar, and Patrick McDonald. "Cosmological parameters from combining the Lyman-alpha forest with CMB, galaxy clustering and SN constraints". In: *JCAP* 0610 (2006), p. 014. DOI: 10.1088/1475-7516/2006/10/014. arXiv: astro-ph/0604335 [astro-ph].

[163] Mark Srednicki. "Entropy and area". In: *Phys. Rev. Lett.* 71 (1993), pp. 666–669. DOI: 10.1103/PhysRevLett.71.666. arXiv: hep-th/9303048 [hep-th].

[164] Alexei A. Starobinsky. "Dynamics of Phase Transition in the New Inflationary Universe Scenario and Generation of Perturbations". In: *Phys. Lett.* 117B (1982), pp. 175–178. DOI: 10.1016/0370-2693(82)90541-X.

[165] Christopher R. Stephens, Gerard 't Hooft, and Bernard F. Whiting. "Black hole evaporation without information loss". In: *Class. Quant. Grav.* 11 (1994), pp. 621–648. DOI: 10.1088/0264-9381/11/3/014. arXiv: gr-qc/9310006 [gr-qc].

[166] L. Susskind and J. Lindesay. *An introduction to black holes, information and the string theory revolution: The holographic universe.* 2005.

[167] Leonard Susskind. "Strings, black holes and Lorentz contraction". In: *Phys. Rev.* D49 (1994), pp. 6606–6611. DOI: 10.1103/PhysRevD.49.6606. arXiv: hep-th/9308139 [hep-th].

[168] Leonard Susskind. "The Anthropic landscape of string theory". In: (2003), pp. 247–266. arXiv: hep-th/0302219 [hep-th].

[169] Leonard Susskind. "The World as a hologram". In: *J. Math. Phys.* 36 (1995), pp. 6377–6396. DOI: 10.1063/1.531249. arXiv: hep-th/9409089 [hep-th].

[170] Leonard Susskind, Larus Thorlacius, and John Uglum. "The Stretched horizon and black hole complementarity". In: *Phys. Rev.* D48 (1993), pp. 3743–3761. DOI: 10.1103/PhysRevD.48.3743. arXiv: hep-th/9306069 [hep-th].

[171] Peter Svrcek and Edward Witten. "Axions In String Theory". In: *JHEP* 06 (2006), p. 051. DOI: 10.1088/1126-6708/2006/06/051. arXiv: hep-th/0605206 [hep-th].

[172] Brian Swingle. "Constructing holographic spacetimes using entanglement renormalization". In: (2012). arXiv: 1209.3304 [hep-th].

[173] Gerard 't Hooft. "Dimensional reduction in quantum gravity". In: *Conf. Proc.* C930308 (1993), pp. 284–296. arXiv: gr-qc/9310026 [gr-qc].

[174] Gerard 't Hooft. "The black hole interpretation of string theory". In: *Nucl. Phys.* B335 (1990), pp. 138–154. DOI: 10.1016/0550-3213(90)90174-C.

[175] Fuminobu Takahashi and Masaki Yamada. "Strongly broken Peccei-Quinn symmetry in the early Universe". In: *JCAP* 1510.10 (2015), p. 010. DOI: 10.1088/1475-7516/2015/10/010. arXiv: 1507.06387 [hep-ph].

[176] Y. Takahashi and H. Umezawa. "Thermo field dynamics". In: *Int. J. Mod. Phys.* B10 (1996), pp. 1755–1805. DOI: 10.1142/S0217979296000817.

[177] Max Tegmark et al. "Dimensionless constants, cosmology and other dark matters". In: *Phys. Rev.* D73 (2006), p. 023505. DOI: 10.1103/PhysRevD.73.023505. arXiv: astro-ph/0511774 [astro-ph].

[178] Michael S. Turner and Frank Wilczek. "Inflationary axion cosmology". In: *Phys. Rev. Lett.* 66 (1991), pp. 5–8. DOI: 10.1103/PhysRevLett.66.5.

[179] W. G. Unruh. "Notes on black hole evaporation". In: *Phys. Rev.* D14 (1976), p. 870. DOI: 10.1103/PhysRevD.14.870.

[180] W. G. Unruh and Robert M. Wald. "Acceleration Radiation and Generalized Second Law of Thermodynamics". In: *Phys. Rev.* D25 (1982), pp. 942–958. DOI: 10.1103/PhysRevD.25.942.

[181] William G. Unruh and Robert M. Wald. "What happens when an accelerating observer detects a Rindler particle". In: *Phys. Rev.* D29 (1984), pp. 1047–1056. DOI: 10.1103/PhysRevD.29.1047.

[182] Mark Van Raamsdonk. "Comments on quantum gravity and entanglement". In: (2009). arXiv: 0907.2939 [hep-th].

[183] Erik Verlinde and Herman Verlinde. "Black Hole Entanglement and Quantum Error Correction". In: *JHEP* 10 (2013), p. 107. DOI: 10.1007/JHEP10(2013)107. arXiv: 1211.6913 [hep-th].

[184] A. Vilenkin and E. P. S. Shellard. *Cosmic Strings and Other Topological Defects*. Cambridge University Press, 2000. ISBN: 9780521654760. URL: http://www.cambridge.org/mw/academic/subjects/physics/theoretical-physics-and-mathematical-physics/cosmic-strings-and-other-topological-defects?format=PB.

[185] Alexander Vilenkin. "The Birth of Inflationary Universes". In: *Phys. Rev.* D27 (1983), p. 2848. DOI: 10.1103/PhysRevD.27.2848.

[186] Robert M. Wald. *General Relativity*. Chicago, USA: Chicago Univ. Pr., 1984. DOI: 10.7208/chicago/9780226870373.001.0001.

[187] Aron C. Wall. "Maximin Surfaces, and the Strong Subadditivity of the Covariant Holographic Entanglement Entropy". In: *Class. Quant. Grav.* 31.22 (2014), p. 225007. DOI: 10.1088/0264-9381/31/22/225007. arXiv: 1211.3494 [hep-th].

[188] J. A. Wheeler. "SUPERSPACE AND THE NATURE OF QUANTUM GEOMETRODYNAMICS". In: (1988).

[189] Edward Witten. "Anti-de Sitter space and holography". In: *Adv. Theor. Math. Phys.* 2 (1998), pp. 253–291. DOI: 10.4310/ATMP.1998.v2.n2.a2. arXiv: hep-th/9802150 [hep-th].

[190] Edward Witten. "Dyons of Charge e theta/2 pi". In: *Phys. Lett.* B86 (1979). [,283(1979)], pp. 283–287. DOI: 10.1016/0370-2693(79)90838-4.

[191] W. K. Wootters and W. H. Zurek. "A single quantum cannot be cloned". In: *Nature* 299 (1982), pp. 802–803. DOI: 10.1038/299802a0.

[192] Matias Zaldarriaga and Uros Seljak. "Gravitational lensing effect on cosmic microwave background polarization". In: *Phys. Rev.* D58 (1998), p. 023003. DOI: 10.1103/PhysRevD.58.023003. arXiv: astro-ph/9803150 [astro-ph].

[193] W. H. Zurek. "Cosmological Experiments in Superfluid Helium?" In: *Nature* 317 (1985), pp. 505–508. DOI: 10.1038/317505a0.

[194] W. H. Zurek. "Entropy Evaporated by a Black Hole". In: *Phys. Rev. Lett.* 49 (1982), pp. 1683–1686. DOI: 10.1103/PhysRevLett.49.1683.