Gravity Informed

Thesis by
Aidan Chatwin-Davies

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

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2018
(Defended May 17, 2018)

Aidan Chatwin-Davies
ORCID: 0000-0003-1406-9271

ACKNOWLEDGEMENTS

In the five years that I have spent as a graduate student at the California Institute of Technology, I have had the privilege to be supported academically, professionally, socially, and emotionally by the finest network of peers, mentors, friends, and family that one could ask for. I would like to recognize the following individuals and groups of people for their help and support:

Co-authors – Ning Bao, Adam Bouland, Charles Cao, Sean Carroll, Nick Hunter-Jones, Adam Jermyn, Achim Kempf, Robert Martin, Jason Pollack, Grant Remmen, and Henry Yuen. We are joined by a bond of punctuation and grammar.

Quantum Spacetime Group Members – Originally known as the Cosmological Quantum Mechanics Group, or just "Cosmo QM" for short, many of the ideas presented in this dissertation had their origins in this group. Here I would like to thank some of the student members of the group:

- Charles Cao – From struggling through homework to braving the frontiers of research together, we have come a long way since we started our first year of studies. Your extraordinary insight into physics never ceases to amaze me, and I am humbled to call you a friend and collaborator. I think you are also the only person with whom I have been to four different continents.

- Tony Bartolotta, Nick Hunter-Jones, Jason Pollack, and Grant Remmen – As part of the old guard, you have been stalwart companions in both research and other matters (particularly concerning Shiganshina, the YITP, Eorzea, and aboard the Artemis, respectively).

- Junyu Liu and Ashmeet Singh – As part of the new guard, I am relying on your youthful brilliance to keep me going now that I am newly old; it has certainly worked to inspire me so far.

Fellow Graduate Students – If graduate students make up the fabric of an institute, then Caltech's fabric is as fine (and durable) as silk. In addition to those students already named, I would like to thank Dávid Guszejnov, Jake Kim, Anna Kómár, Greg MacCabe, Christopher White, and Minyoung You for their friendship and support.

The "Mates" – I would like to thank Alex Turzillo (officemate), with whom it was an absolute pleasure to share an office for both working hours and after-hours shenanigans; Tomas Jochym-O'Connor (housemate), with whom I enjoyed bonding over physics, French, and of course our favourite mustache-sporting Italian plumber; and Aleksander Kubica (housemate), with whom I shared all aspects of graduate student life during my first four years at Caltech.

Dear Friends – Rayner Andersen, Wilson Brenna, John Churchill, Stephen Favron James Hawyard, and Scott Jakes. Having known them for over a decade, I can always turn to these friends for candid counsel, relentless camaraderie, and unshakable support. In particular, I would like to recognize Stephen for continuing to host me for extended periods of time when I visit Waterloo, and John for being an enduring companion on adventures in both real and virtual worlds.

Family – I would like to thank my mother, Ann Chatwin; my father, Ian Davies; my grandmother, Loie Chatwin; and my sister, Robyn Chatwin-Davies, for always giving me a place that I can call home.

Mentors – The following individuals have helped define me as a physicist:

- Gil Refael – Assisting with Physics 50 under your guidance has been the best teaching assistant job that I could have hoped for. Participating in this class has helped me learn how to think like a physicist just as much as any other form of academic training.

- Achim Kempf – Having transitioned from master's advisor to collaborator, I continue to value your sage advice on all matters relating to life as a physicist. I am particularly grateful for the generous support that you provide to host me for visits at the University of Waterloo.

- Ning Bao – You arrived at Caltech at a critical point of my graduate career; your ability to support, advise, and empathize with me went above and beyond any expectation of a postdoctoral scholar. I look forward to continuing to work together as both close collaborators and close friends.

Dissertation and Candidacy Committee Members – Sean Carroll, Cliff Cheung, John Preskill, and Alan Weinstein. Their thoughtful guidance led to significant improvements in the direction and content of this dissertation.

Advisor – Sean Carroll.

The quality of graduate student advising can vary wildly, even at an institution as reputable as Caltech. Fortunately for me, the experience I had with you as an advisor sits at the highest of echelons in terms of graduate advising. You were always available to talk about research, academic duties, and life in general—and you were surprisingly tolerant of my tendency to barge into your office unannounced for short physics discussions! Working with you and witnessing your unwavering commitment to integrity and excellence first-hand has inspired me to contribute to the physics community to the best of my abilities and to engage with the full cross-section of its members, from experts, to students, to members of the public. Thank you.

ABSTRACT

Formulating a universally satisfactory theory of quantum gravity is a long-standing open problem in theoretical physics. Relatively recently, the use of techniques from quantum information has emerged as a powerful tool for analyzing phenomena that lie at the intersection of quantum theory and gravitation. This thesis describes several advances and novel proposals that were made regarding information theoretic aspects of quantum gravity in three broad areas: holography, cosmology, and the black hole information problem.

Regarding holography, we first assess the differences between typical holographic states and fully random states. Next, we show that determining Ryu-Takayanagi surfaces in $\mathrm{AdS}_3/\mathrm{CFT}_2$ is computationally easy from a complexity-theoretic standpoint. Finally, we identify precise consistency conditions that constrain the validity of an early tensor network model for the AdS/CFT correspondence that uses the Multiscale Entanglement Renormalization Ansatz (MERA).

Regarding cosmology, we propose an alternative interpretation of the MERA as a discretization of de Sitter spacetime. Next, we return to holographic ideas and show that an appropriately-defined Generalized Second Law implies a cosmic no-hair theorem for certain classes of cosmological spacetimes. Finally, we advance an information-theoretic proposal for calculating the signature of a quantum gravity-motivated, fully covariant, natural ultraviolet cutoff in the spectrum of inflationary perturbations.

Regarding the black hole information problem, we begin by exhibiting a simple protocol which, under highly specific circumstances, allows one to retrieve a single qubit from a black hole. Next, we propose an operational resolution of the black hole information problem in which observers who enter the black hole could never detect an inconsistency between their experiences and quantum mechanics due to the finite amount of time available before reaching the central singularity. Finally, we discuss a proposal to understand the emergence of an ensemble of definite geometries

during the process of black hole evaporation as a decoherence process, as well as its implications for the black hole information problem.

PUBLISHED CONTENT AND CONTRIBUTIONS

The following articles constitute Chs. 2-7 and 9-11 of this dissertation. I was a major contributing author for all of these articles and participated in all parts of the writing process.

[1] N. Bao and A. Chatwin-Davies, "Puzzles and pitfalls involving Haar-typicality in holography," arXiv:1708.08561.

[2] N. Bao and A. Chatwin-Davies, "The complexity of identifying Ryu-Takayanagi surfaces in AdS3/CFT2," *JHEP* **11** (2016) 034, arXiv:1609.01727.

[3] N. Bao, C. Cao, S. M. Carroll, A. Chatwin-Davies, N. Hunter-Jones, J. Pollack, and G. N. Remmen, "Consistency conditions for an AdS multiscale entanglement renormalization ansatz correspondence," *Phys. Rev. D* **91** (2015) 125036, arXiv:1504.06632.

[4] N. Bao, C. Cao, S. M. Carroll, and A. Chatwin-Davies, "De Sitter space as a tensor network: Cosmic no-hair, complementarity, and complexity," *Phys. Rev. D* **96** (2017) 123536, arXiv:1709.03513.

[5] S. M. Carroll and A. Chatwin-Davies, "Cosmic equilibration: A holographic no-hair theorem from the generalized second law," *Phys. Rev. D* **97** (2018) 046012, arXiv:1703.09241.

[6] A. Chatwin-Davies, A. Kempf, and R. T. W. Martin, "Natural covariant Planck scale cutoffs and the cosmic microwave background spectrum," *Phys. Rev. Lett.* **119** (2017) 031301, arXiv:1612.06445.

[7] A. Chatwin-Davies, A. S. Jermyn, and S. M. Carroll, "How to recover a qubit that has fallen into a black hole," *Phys. Rev. Lett.* **115** (2015) 261302, arXiv:1507.03592.

[8] N. Bao, A. Bouland, A. Chatwin-Davies, J. Pollack, and H. Yuen, "Rescuing complementarity with little drama," *JHEP* **12** (2016) 026, arXiv:1607.05141.

[9] N. Bao, S. M. Carroll, A. Chatwin-Davies, J. Pollack, and G. N. Remmen, "Branches of the black hole wave function need not contain firewalls," arXiv:1712.04955.

TABLE OF CONTENTS

Part III: The Black Hole Information Problem

LIST OF ILLUSTRATIONS

Preliminaries

Chapter 1

INTRODUCTION

The two most important physical theories of the last century are general relativity and quantum mechanics. It is no exaggeration to say that every major development in physics has relied on one or both of these two theories since their mainstream adoption, and an important pending development is the satisfactory characterization of their direct intersection. Finding a successful theory of quantum gravity that is universal in scope is a major open direction of research.

Of course, given that we have successful quantum descriptions of physical phenomena as well as successful general relativistic descriptions of gravitational phenomena, it is natural to hunt for a quantum description of gravity. Besides logical satisfaction, however, practical reasons require us to confront the issue of quantum gravity.

By now there is ample evidence, both direct and indirect, that there exist energy scales and regimes of curvature in the universe for which both quantum and gravitational effects are important—the two most conspicuous being the very early universe and black holes, due to the extreme energy and curvature scales involved. Less conspicuous settings include, for example, satellite-based quantum communication experiments, where spacetime curvature influences the distribution and exploitation of entanglement among parties which sit at different points in the Earth's gravitational field [10]. Another setting is the conventional laboratory, where certain precision tabletop experiments propose to measure interference between paths which cross different gravitational potentials [11], while others propose to place mesoscopic objects in superpositions of different states of motion [12].

There are myriad approaches to trying to quantize gravity (or, alternatively, to trying to gravitize quantum theory[1]). Among all of these approaches, an important

[1] I attribute this terminology to ChunJun Cao. I first heard it in his Ph.D. thesis defence on May 1, 2018 at the California Institute of Technology.

tool to emerge over the last few decades is the use of quantum information techniques as a means of analyzing quantum gravity. This is because the abstractions of information theory appear to be a common language between gravitational and quantum mechanical systems.

A linchpin of this program is the link between entanglement entropy and the geometric area of surfaces in spacetime. Thought experiments in relativity indicate that the entropy of gravitational systems is bounded by the area of appropriate bounding surfaces [13, 14]. If spacetime is to admit a description in terms of quantum degrees of freedom in a Hilbert space, then this bound suggests that gravity is *holographic*. That is, the number of degrees of freedom corresponding to a region of spacetime should not scale with the volume of the region, but rather with the area of a bounding surface in one lower dimension. The entanglement entropy of these degrees of freedom is then the source of the entropy of the gravitational system.

An important testing ground that explicitly realizes holography and which further purports to be an exact theory of quantum gravity is the AdS/CFT correspondence [15, 16]. According to AdS/CFT, the states of certain conformal field theories (CFTs) are in exact correspondence with states of a quantum gravitational theory of asymptotically Anti de Sitter (AdS) spacetimes in one higher dimension, where the CFT may be identified with the spacetime boundary. Within AdS/CFT, the link between entropy and area is made explicit by the Ryu-Takayanagi formula for static spacetimes in the large-N limit [17], which relates the entanglement entropy of a region A on the boundary to the area of a surface $\tilde{A}$ in the bulk:

$$S(A) = \frac{\text{area}(\tilde{A})}{4G}. \tag{1.1}$$

The bulk surface $\tilde{A}$ is chosen such that $\partial A = \partial \tilde{A}$, $\tilde{A}$ is homologous to A, and such that it has maximal area over all surfaces that satisfy the first two criteria.

The first part of this dissertation deals with topics in AdS/CFT. Ch. 2 gives a pedagogical account of what it means to be a typical holographic state. Not all states in AdS/CFT have well-defined semiclassical geometric duals, and so one must

be careful when trying to derive statistics on average holographic states.

Next, Ch. 3 is directly concerned with the Ryu-Takayanagi formula itself. This chapter assesses the computational complexity of computing Ryu-Takayanagi surfaces in AdS_3/CFT_2. This task is computationally difficult in higher dimensions, but the task is determined to be computationally easy in the special case of AdS_3/CFT_2 where the geometry is particularly simple. As a result, combinatorial complexity is clearly separated from geometric considerations.

The next chapter broaches a technique that has important applications to AdS/CFT: tensor networks. These were originally developed for use in condensed matter theory as a way to efficiently represent complicated quantum states of many-body systems. In essence, tensor networks trade algebraic complexity for geometric information, and as such have been adopted by the gravity community as possible "building blocks" for models of spacetime. Ch. 4 examines the limitations of a model for AdS/CFT built out of the Multiscale Entanglement Renormalization Ansatz (MERA) tensor network.

Ch. 5 continues with the theme of tensor network models for spacetime while opening the second part of this dissertation, which pertains to topics in cosmology. As an alternative to AdS/CFT, this chapter proposes an interpretation for the MERA as a cosmological de Sitter spacetime.

Ch. 6 returns to holographic ideas, but now in a cosmological context. In particular, a generalized second law of thermodynamics which relates the increase in entropy of gravitational systems to the increase in area of families of compact surfaces in the geometry is examined. Here it is shown that this generalized second law implies a cosmic no-hair theorem. That is, it is established that the increase of generalized entropy up to a finite maximum value in certain cosmological models implies that the spacetime asymptotes to a de Sitter configuration in the future. In this sense, de Sitter is a cosmological equilibrium state.

Then, Ch. 7 ends the cosmological part of this dissertation with a topic that is

situated in a very different regime of cosmology: the inflationary universe. This chapter is concerned with characterizing the signatures of Planck-scale physics in the spectrum of inflationary perturbations. More precisely, this chapter presents the computation of the signature of a natural covariant minimum length scale, where the minimum length scale takes the form of a covariant bandlimit in the sense of Shannon information theory. In other words, the existence of a finite natural minimum length scale is reflected as a cutoff on the density of degrees of freedom in spacetime.

The last part of this dissertation covers topics having to do with the other major area of overlap between quantum theory and general relativity mentioned at the beginning of the introduction: the black hole information problem. Here, the application of information theory to black holes leads to a stark dilemma: it would appear that black holes either destroy quantum information, or local effective field theory must break down in a significant way in the presence of gravity. This dilemma is known as the black hole information problem and is further reviewed in Ch. 8.

Ch. 9 constitutes a first foray into addressing the black hole information problem, in which an information-theoretic protocol is discussed whereby it is possible to retrieve a single qubit from a black hole under special conditions. However, the seriousness of the black hole information problem lies in longer timescales, over which a significant portion of the black hole evaporates.

Two approaches to overcoming the black hole information problem on longer timescales are discussed in the next two chapters. Ch. 10 explores an operational resolution. The basic idea is that an observer who crosses the event horizon of a black hole does not have enough time to detect violations of quantum mechanics before reaching the singularity, under reasonable assumptions about their computing power.

The spirit of Ch. 11 is to take seriously the idea that, on timescales of order the black hole mass, evaporation does not lead to a single black hole geometry, but rather an ensemble of possible definite geometric configurations. Decoherence is proposed as a mechanism for understanding the quantum mechanical emergence of this ensemble. Given that the black hole information problem is based on postulates that are

partially formulated using effective field theory on a curved spacetime background, said postulates are re-examined in light of the decoherence proposal to conclude that there may be no problem in the first place. An important independent lesson from this chapter is that foundational aspects of quantum theory appear to have important ramifications for quantum gravitational phenomena.

1.1 How it all fits together

As an alternative way of understanding how the topics addressed in this thesis fit into the bigger picture of quantum gravity, I have laid them out in a mind map alongside other principles, ideas, and directions in Fig. 1.1. A short glossary of the terms appearing in this mind map are also given below, along with selected references. The selection of topics shown is by no means comprehensive, but instead reflects my own interests and those of my colleagues.

Glossary

AdS/CFT – The Anti de Sitter/Conformal Field Theory correspondence [15, 16].

Amplitudes – The program of calculating scattering amplitudes for processes that involve gravitons and other species in the framework of perturbative quantum gravity.

AMPS – Almheiri, Marolf, Polchinski, and Sully; authors of the celebrated "firewalls" papers which succinctly formulated the modern black hole information problem [18, 19].

Black Holes/Information – By this, I mean the specific sub-field of black hole physics that examines the issue of whether the evolution of black holes is unitary or not.

Bulk Emergent Gravity – The proposal to understand the emergence of a spacetime metric from quantum correlations in a qudit system and the emergence of gravitational dynamics from the dynamics of entanglement [20, 21].

Classicalizaton/Decoherence – The process by which manifestly quantum fluctua-

tions of the inflaton field become classical stress-energy/curvature perturbations. See, e.g., [22].

Complementarity – The idea that an infalling observer and an observer who remains outside of a black hole can have complementary descriptions of where they think quantum information is stored in spacetime [23].

Distance/Mutual Info – This refers to Van Raamsdonk's finding that mutual information in the CFT can serve as a proxy for geodesic spacetime distance in AdS [24].

Emergence – By this I mean any "bottom-up" approach to quantum gravity, which starts with a quantum mechanical systems and aims to discover gravity lurking within.

Entropy Cone – Constraints on holographic states from saturating multi-party entanglement entropy inequalities [25].

Error Correction – Seminal references on the interpretation of AdS/CFT as error correction are Refs. [26, 27].

Extremal Surface Zoo – The specific program of interpreting extremal bulk surfaces as entropic quantities, e.g., Ref. [28].

Generalized Holography – The program of realizing holography in other gravitational theories besides those that describe asymptotically AdS spacetimes.

Generalized Second Law – See Ref. [29].

Geometrization – By this I mean the program of interpreting information theoretic objects in the CFT as geometric structures in the gravitational bulk.

Gravitational Wave Spectroscopy – The ambition to eventually extract detailed information about the quantum structure of black holes and the epochs which precede recombination from gravitational radiation [30].

Holographic Codes – Quantum error correcting codes that serve as toy models for holography, e.g., Ref. [31].

Holography – The idea that it may be possible to encode complete information about a gravitating system in one fewer dimension than is apparent in the gravitational description.

HRT – Hubeny-Rangamani-Takayanagi; a covariant version of the Ryu-Takayanagi formula [32].

Perturbative Quantum Gravity – A regime of quantum gravity where gravitational perturbations are treated as quantized propagating excitations on a background spacetime.

Planck-Scale Effects in Cosmology – The program of looking for the signatures of Planck-scale effects in cosmic microwave background (CMB) data.

Quantum Energy Conditions – Generalizations of classical stress-energy conditions that aim to retain their validity at next to leading order in $1/G$ in gravity [33, 34].

Quantum Mereology – An attempt to understand the emergence of preferred factorizations of Hilbert space on entropic grounds [35].

Ryu-Takayanagi – See Ref. [17].

Semiclassical Quantum Gravity – By this I mean approaches and topics in quantum gravity in which a semiclassical notion of a background spacetime still figures prominently.

State Measures – Refers to the general topic of characterizing what proportion of CFT states have well-defined geometric duals, as well as the statistics of this distribution.

Tensor Networks – For a review, see Ref. [36].

Weak Gravity Conjecture – The conjecture that gravity coupled to a species with a $U(1)$ charge should have state such that its charge exceeds its mass. See, e.g., Refs. [37, 38].

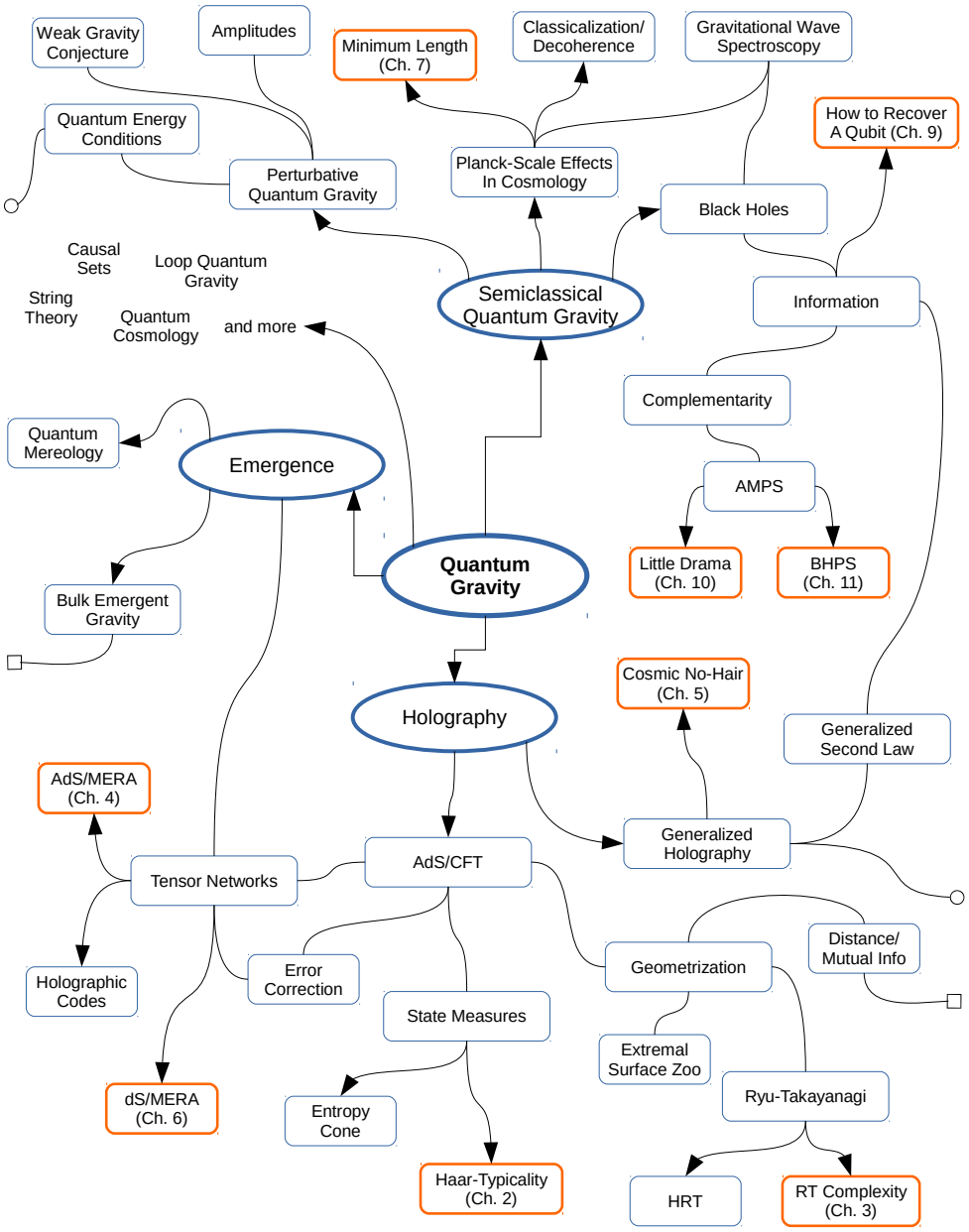

Figure 1.1: Approaches to quantum gravity, with emphasis on approaches that have captured the interest of members of the quantum spacetime group at the California Institute of Technology, as well as topics discussed in this dissertation.

Part I

Holography

Chapter 2

PUZZLES AND PITFALLS INVOLVING HAAR-TYPICALITY IN HOLOGRAPHY

Typical holographic states that have a well-defined geometric dual in AdS/CFT are not typical with respect to the Haar measure. As such, trying to apply principles and lessons from Haar-random ensembles of states to holographic states can lead to apparent puzzles and contradictions. We point out a handful of these pitfalls.

This chapter is available in preprint form as Ref. [1], N. Bao and A. Chatwin-Davies, "Puzzles and pitfalls involving Haar-typicality in holography," `arXiv:1708.08561`.

2.1 Introduction

Typical states and the technique of averaging over ensembles of states are powerful tools in quantum information theory. In high-energy physics, these tools have been a central component of quantum gravity studies over the last several years. In particular, the notion of Haar-typical states and Page's theorem are two workhorses of quantum information in quantum gravity.

Roughly speaking, a Haar-typical state is a fully-random quantum state. Given a finite-dimensional Hilbert space $\mathcal{H}$ and any reference state $|\psi_0\rangle \in \mathcal{H}$, a Haar-typical state may be thought of as a realization of the random variable $|\psi(U)\rangle = U|\psi_0\rangle$, where U is a random unitary matrix drawn with uniform probability from the set of all unitary matrices. The uniform probability distribution over the unitaries is the normalized Haar measure over the set of unitary matrices. For convenience, we will just refer to this as "the Haar measure" for the rest of this note.

The Haar measure is a key ingredient in both the statement and proof of Page's theorem [39]. Suppose now that the Hilbert space splits into two subfactors, $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$. If $|\psi\rangle$ is a Haar-typical state on $\mathcal{H}$, then Page's theorem essentially says that, with high probability, the reduced state of $|\psi\rangle$ in the $\mathcal{H}_A$ subfactor is very

nearly maximally entangled with the part of the state in the $\mathcal{H}_B$ subfactor if the dimension of $\mathcal{H}_A$ is much smaller than that of $\mathcal{H}_B$. Taken together, Page's theorem and Haar typicality are a precise statement of the notion that a small subsystem generically tends to be maximally entangled with its larger complement in a fixed Hilbert space.

Page's theorem has been fruitfully applied to the physics of semiclassical black holes [18, 40–42]. Here, one usually considers a collection of matter that is initially in a pure state and that collapses into a black hole, which over time dissipates via the process of Hawking evaporation [43]. From a quantum-mechanical standpoint, the whole process is modelled as taking place in a single Hilbert space, $\mathcal{H}$, in which a finite number of degrees of freedom are divided up between the black hole and radiation. Of course, the factorization of $\mathcal{H}$ varies from Cauchy slice to Cauchy slice as the black hole evaporates, but if black holes do not destroy information, then the evolution of the total state is unitary and the size of the total Hilbert space must be constant in this model.

While the quantum-gravitational dynamics of black hole degrees of freedom are unknown, for information-theoretic purposes at least, it seems that it is reasonable to model the dynamics in the black hole's Hilbert space by Haar-random unitary evolution (or a 2-design approximation thereof [42]), at least on timescales that are shorter than the rate of Hawking emissions so that the size of the black hole Hilbert space is constant. Therefore, shortly after the black hole forms[1], the total state in $\mathcal{H}$ is Haar-typical in this model, and so Page's theorem may be used to study the entanglement properties of the state across the changing factorization of $\mathcal{H}$ during the subsequent evolution. For example, this sort of analysis revealed that if you toss a small quantum system into a black hole, then information about its state is rapidly returned via Hawking radiation once the black hole has given up more than half of its degrees of freedom through evaporation [42]. Similar considerations are at the heart of the ongoing debates that surround complementarity [23], the black

[1] Or more precisely, one scrambling time after the black hole forms [42].

hole information problem [44], and the firewall puzzle [18].

However, an area in which Page's theorem and Haar-typicality have significantly less applicability is AdS/CFT [15, 16]. This is because a typical holographic state in a large-N CFT that has a well-defined geometric gravitational dual is not typical with respect to the Haar measure. To see this, note that holographic states are significantly less entangled on small scales, per the Ryu-Takayanagi formula, than Haar-typical states. Given some small spacelike region A in the CFT, the reduced state on A of a Haar-typical state per Page's theorem must have extensive entanglement entropy that scales like the volume of A. For a holographic state, when A is small enough, the minimal surface anchored to ∂A only probes the region of the bulk gravitational dual near its asymptotically AdS boundary. Therefore, the area of the minimal surface, and hence the entropy of the reduced state on A, must be sub-extensive in the volume of A. Haar-typical states are therefore not holographic. Nevertheless, careful use of random state statistics and suitable generalizations of Haar-typicality, such as typicality with respect to the microcanonical ensemble in a fixed energy window, can serve as useful tools for holography.

That typical holographic states are not typical with respect to the Haar measure is well-known in the community. Nevertheless, it is quite easy to momentarily overlook this fact, which can lead to apparent puzzles in holography. Our main goals in writing this short note are to highlight the fact that Haar-typical states are not holographic and to point out a handful of potential pitfalls that could arise from overextending Page's theorem when it enters into discussions about holography in the literature. We hope that both veterans and novices will find this note to be digestible and helpful in avoiding confusions that have befallen us in the past.

In Section 2.2, we review the precise definition of Haar-typicality and the precise statement of Page's theorem, and we reiterate the argument for why Haar-typical states are not holographic in commensurate language. Then, in Section 2.3, we point out several puzzles and potential pitfalls in the literature. We offer some concluding remarks in Section 2.4.

2.2 Properties of random states

Let $\mathcal{H}$ be a finite-dimensional Hilbert space with dimension n and consider the group of unitary transformations, $\mathrm{U}(n)$, acting on $\mathcal{H}$.

Definition 2.2.1. *The normalized* Haar measure *on* $\mathrm{U}(n)$ *is the unique measure* μ *measuring subsets of* $\mathrm{U}(n)$ *such that:*

- $\mu(U\mathcal{S}) = \mu(\mathcal{S}U) = \mu(\mathcal{S})$ *for all* $\mathcal{S} \subset \mathrm{U}(n)$ *and* $U \in \mathrm{U}(n)$*, where* $U\mathcal{S} = \{UV \mid V \in \mathcal{S}\}$ *and* $\mathcal{S}U$ *is similarly defined.*
- $\mu(\mathrm{U}(n)) = 1$

From the properties above, it also follows that μ is non-negative, so the normalized Haar measure on $\mathrm{U}(n)$ is a uniform probability distribution on the group of unitary transformations as such.

To state Page's theorem, we also need the trace norm, which is defined as $\|T\|_1 = \mathrm{tr}\,\sqrt{T^\dagger T}$ for any linear operator T on $\mathcal{H}$. The trace norm gives a good notion of distinctness of states because if $\|\rho - \sigma\|_1 < \epsilon$ for any two density operators (i.e., states) ρ and σ on $\mathcal{H}$, then $\|P(\rho - \sigma)\|_1 < \epsilon$ for any projector P so that the probabilities for measurement outcomes in the states ρ and σ are close [45].

Now suppose that $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$, where $\dim \mathcal{H}_A = d_A$ and $\dim \mathcal{H}_B = d_B$ with $n = d_A d_B$. Without loss of generality, suppose that $d_A \leq d_B$. A precise statement of Page's theorem is as follows[2]:

Theorem 2.2.2 (Page)**.** *Let* $|\psi_0\rangle \in \mathcal{H}$ *be a fixed reference state and let* $|\psi(U)\rangle = U|\psi_0\rangle$ *for any* $U \in \mathrm{U}(n)$*. Let* $\rho_A(U) = \mathrm{tr}_B\,|\psi(U)\rangle\langle\psi(U)|$ *denote the reduced state of* $|\psi(U)\rangle$ *on* $\mathcal{H}_A$*. Then it follows that*

$$\int d\mu(U)\ \left\| \rho_A(U) - \frac{I_A}{d_A} \right\|_1 \leq \sqrt{\frac{d_A}{d_B}}, \tag{2.1}$$

[2] As a historical note, the theorem originally proved by Page [39] was formulated in terms of entanglement entropies and built upon earlier works by Lubkin [46] and Lloyd & Pagels [47]. The version of Page's theorem given here is based on the statement appearing in [45]. A detailed yet accessible proof of Page's theorem can be found in Sec. 10.9 of [48] under the name of "the decoupling inequality."

where I_A is the identity operator on $\mathcal{H}_A$.

Page's theorem therefore says that the *average* distance between a random reduced state $\rho_A(U)$ and the maximally mixed state on $\mathcal{H}_A$, I_A/d_A, is bounded by $\sqrt{d_A/d_B}$. The bite of Page's theorem comes when $d_A \ll d_B$, in which case the average distance is very small. We note this can be the case for qubit systems even when the number of qubits associated with A is only a little smaller than those associated with B, as the dimensionality scales exponentially in the number of qubits. In this case, the interpretation of Page's theorem is that the reduced state on a small subfactor of a randomly-chosen pure state is, with high probability, close to the maximally mixed state. Or, in other words, for a Haar-typical state, the reduced state on a small subfactor of Hilbert space is very nearly maximally entangled with its complementary degrees of freedom.

Now consider a CFT in D dimensions. Let the theory have both ultraviolet and infrared regulators so that it sits on a periodic lattice with spacing ϵ with (finite-dimensional) Hilbert space subfactors at every vertex, and let $\mathcal{H}$ be the Hilbert space of the regulated theory. For example, the theory could describe an Ising spin chain, where the CFT is recovered in the continuum limit. It is formally necessary to work with a finite-dimensional regularization so that the Haar measure and Page's theorem apply. Consider a subregion of the CFT labelled by A that has a characteristic linear dimension l, such as a $(D-1)$-dimensional ball. Now let $|\psi\rangle$ be some Haar-typical state on $\mathcal{H}$. Then, provided that A is a small subregion of the entire CFT, the reduced state on A, call it ρ_A, is very close to being maximally mixed by Page's theorem. In particular, this means that its von Neumann entropy scales extensively with the volume of A:

$$S(\rho_A) \propto \left(\frac{l}{\epsilon}\right)^{D-1}. \tag{2.2}$$

The reason why a Haar-typical state is not holographic is because the von Neumann entropy in Eq. (2.2) above is in tension with the Ryu-Takayanagi formula [17]. Suppose now that the CFT in question has a state $|\phi\rangle$ with a well-defined, asymptotically

AdS geometric dual in $D+1$ dimensions where we can think of the CFT as living on the bulk geometry's boundary. Per the Ryu-Takayanagi formula, the von Neumann entropy of the reduced state on A, call it σ_A, is given by the area in Planck units of a bulk minimal surface $\tilde{A}$ that is homologous to A and such that $\partial A = \partial \tilde{A}$:

$$S(\sigma_A) = \frac{\text{area}(\tilde{A})}{4G}. \tag{2.3}$$

The tension comes when A is small enough such that the minimal surface $\tilde{A}$ only probes the near-boundary region, which has asymptotic AdS geometry. Since such $\tilde{A}$ only sees AdS geometry, its area will be subextensive compared to the volume of A in the boundary. For example, when $D = 3$ and when A is a small disk of radius l, the area of $\tilde{A}$ scales like l/ϵ, cf. Eq. (2.2). In general, the scaling is

$$\text{area}(\tilde{A}) \propto \int_{\epsilon/l}^{1} d\zeta \; \frac{(1-\zeta^2)^{(D-3)/2}}{\zeta^{D-1}} \leq \int_{\epsilon/l}^{1} \frac{d\zeta}{\zeta^{D-1}}, \tag{2.4}$$

which is subextensive compared to $(l/\epsilon)^{D-1}$. Therefore, typical holographic states cannot be typical with respect to the Haar measure, and arguments which depend critically on Haar typicality will generally not apply to states with classical holographic bulk duals.

2.3 Puzzles resolved and pitfalls espied

Having reviewed Haar-typicality and Page's theorem, we now identify a handful of situations in the literature on holography where intuition from and use of Haar-typical states can be misleading. We also discuss two situations in which the use of Haar-typicality is appropriate.

2.3.1 Measures of holographic states

A current area of research is what fraction of quantum states are allowed to be holographic states. In particular, the measure computed using the entropy cone [25] seems to conflict with a recent numerical study by Rangamani and Rota [49] for reasons that we now clarify.

In their work, Rangamani and Rota study how measures of entanglement and the structure of entanglement among different partitions of a pure state characterize the structure of the state. Given a randomly-chosen pure state of N qubits, they first trace out $k < N$ qubits and then compute various measures of entanglement among partitions of the remaining $N - k$ qubits. For example, one entanglement measure that they check is whether states generated in this way obey monogamy of mutual information (MMI):

$$S_{AB} + S_{BC} + S_{AC} \geq S_A + S_B + S_C + S_{ABC}. \tag{2.5}$$

In the above, A, B, and C denote subsets of the remaining $N - k$ qubits, and all permutations such that $3 \leq |ABC| \leq N-k$ are checked. ($|A|$, $|AB|$, etc. denotes the number of qubits in A, AB, etc.) MMI is an inequality which holographic states must obey but generic quantum states need not [50]. In their work, it was found that when the method described above is Monte Carlo iterated, almost all states generated in this way satisfy MMI. It is further believed that the higher party-number holographic inequalities [25] would be generically obeyed, as well.

The entropy cone measure, by contrast, does not randomly generate states directly, but rather the entanglement entropies $S_A, S_B, S_{AB}, \ldots$ directly. Once the requisite entanglement entropies are generated, it is then determined whether they (a) are entanglement entropies that are valid for a quantum state and (b) satisfy the further holographic entanglement entropy inequalities, such as MMI. The ratio of the number of sets of randomly generated entanglement entropies that are consistent with both holography and quantum mechanics as a fraction of the number of such entropies consistent with quantum mechanics is then computed. When using this measure, it is found that just over half of all sets of entropies consistent with quantum constraints do not obey the holographic inequalities for three parties, and that this fraction appears to fall off rapidly as a function of party number [51].

We stress that Rangamani and Rota's goal was not to compute a measure of holographic states; however, the entropy cone measure is somewhat puzzling in light of how weakly-constraining MMI is in their numerical assays. At this point, it is

useful to note that Rangamani and Rota's construction is generic with respect to the Haar measure; the way that the states are constructed there could have been equivalently done via the application of a Haar-random unitary. For example, from the perspective of this measure, if $|A|, |B|, |C| \ll N - k$, then generically we would expect MMI to be not only satisfied, but saturated—not because of any holographic consideration, but because MMI is a balanced inequality and all of the entropies are approximately equal to the logarithms of the dimensions of the states' selfsame Hilbert spaces by Page's theorem.

In general, MMI tends to be satisfied for arbitrary partitions of Haar-typical states, since $S_K = S_{K^c}$ with Page's theorem applied to the complement K^c for any collection of qubits such that $|K| > N/2$. If holographic states are fundamentally atypical with respect to the Haar measure, then attempting to conclude which fraction of states can be holographic using a measure that is typical with respect to the Haar measure would yield false positives. The entropy cone measure, by contrast, is unbiased with respect to subregion entropies, as it does not directly generate the states.

2.3.2 Error correction

The assumption that holographic states are well-governed by Page's theorem also appears within the original paper positing the connection between quantum error correction and AdS/CFT [26]. Here, typicality with respect to the Haar measure is used to argue that deletion of arbitrary sets of l qubits in the boundary can be totally corrected given that more than half of the boundary conformal field theory is retained. The reason given for this sharp transition is that this is the point at which the deleted portion of the CFT changes between being just less than and just more than half of the number of qubits in the CFT. Then, once the deleted region becomes less than half of the CFT, it becomes exponentially close to maximally mixed. In particular, this is how the relationship between the error correction picture and the entanglement wedge was originally argued. For a more detailed and complete picture of this argument, see [26].

In reality, the constraint provided by Ryu-Takayanagi prevents the deleted region of the boundary CFT from being close to maximally mixed, and thus a key portion of the argument above is no longer supported. Indeed, while it is true that a generic state with respect to the Haar measure can be corrected by a typical code with a random k qubit code subspace of n qubits so long as $n - 2l - k \gg 1$, because holographic states are atypical with respect to the Haar measure, this statement has little traction in holography. Thus, this portion of the error correcting picture of holography is not actually suggestive of the entanglement wedge, unless another measure that holographic states can be typical with respect to can be shown to yield similar results. Further works [27,52] eventually established the relationship between the entanglement wedge and quantum error correction. Nevertheless, we stress that these Page-type arguments about the relationship between quantum error correction and holograph are specious, and must be taken with a large grain of salt, if at all, particularly in the construction of new arguments in this field.

2.3.3 Butterfly effects and shockwaves

Haar-randomness has also been used in the context of [53] to model the behavior of a shockwave acting on the left half of a thermofield double state. To the past of the shockwave, the thermofield double state has a classical bulk black hole geometry. The state is conjectured to have a classical bulk geometry after the shockwave, whose sole effect should be the displacement of the event horizon. This would be difficult to realize by approximating the shockwave as a Haar-random unitary. Based on the previous discussion, acting with a Haar random unitary on a CFT state with a well-defined geometric dual would typically create a generic state whose dual geometry (at least on the left half) would be totally non-classical (in order for boundary subregions to obey Page's theorem). This is undesirable for approximating shockwave geometries, as one loses the ability to reproduce holographic entanglement entropies on spacelike slices of the left region.

2.3.4 Random tensor networks

We note, however, that holographic models can exploit Haar-randomness, as long as the objects that are Haar-random are Haar-random below the curvature scale. For example, the random tensor model of holography [54] has Haar-random unitaries involved in the creation of each individual tensor site. This should not, however, create a global Haar-typical state, as the spatial arrangement of the tensors in the tensor network would provide a structure to the entanglement that is consistent with that of holography. In this way the overall global state avoids being Haar-typical. It is possible that such models may have issues with reproducing physics below the curvature scale, but that is outside of the scope of this work.

2.3.5 Other probability distributions

Similarly, states that are typical with respect to other random state distributions that deviate from the Haar measure on the full Hilbert space can also serve as good models for holographic states. For example, let $\mathcal{H}_{[E,E+\delta E]}$ denote the subspace spanned by the eigenstates of the Hamiltonian whose energies lie within the range $[E, E+\delta E]$ for some $\delta E \ll E$. A random state drawn with uniform probability from this energy shell (or in other words, sampled from the microcanonical ensemble) typically has a pure-state black hole dual [55–57]. In this case, the Haar measure on $\mathcal{H}_{[E,E+\delta E]}$ is admissible because the resulting measure on $\mathcal{H}$ is not itself the Haar measure.

2.4 Conclusion

A randomly-chosen state is probably not holographic. We all know this fact very well, but it can be easy to overlook when tempted with results from Haar-typical states. From this perspective, we clarified some potentially confusing points in the literature. We hope that this will help newcomers to the field avoid being misled by Haar-random intuition.

Acknowledgements

We would like to thank Hirosi Ooguri, Daniel Harlow, Massimiliano Rota and Mukund Rangamani for discussions. We would especially like to thank Bogdan Stoica, Sam Blitz, and Veronika Hubeny for collaboration in the early part of this work. N.B. was supported in part by the DuBridge Postdoctoral Fellowship, by the Institute for Quantum Information and Matter, an NSF Physics Frontiers Center (NFS Grant PHY-1125565) with support of the Gordon and Betty Moore Foundation (GBMF-12500028). A.C.-D. was supported by the Gordon and Betty Moore Foundation through Grant 776 to the Caltech Moore Center for Theoretical Cosmology and Physics. This work was supported by the U.S. Department of Energy, Office of Science, Office of High Energy Physics, under Award Number DE-SC0011632.

Chapter 3

THE COMPLEXITY OF IDENTIFYING RYU-TAKAYANAGI SURFACES IN ADS_3 / CFT_2

We present a constructive algorithm for finding Ryu-Takayanagi surfaces in AdS_3/ CFT_2 which exploits previously noted connections between holographic entanglement entropy and max-flow/min-cut. We then show that the algorithm runs in polynomial time.

This chapter was published as Ref. [2], N. Bao and A. Chatwin-Davies, "The complexity of identifying Ryu-Takayanagi surfaces in AdS3/CFT2," JHEP ***11*** *(2016) 034,* `arXiv:1609.01727`.

3.1 Introduction

The calculation of entanglement entropy S is a key aspect in understanding the degree of quantumness of a system. While this is a problem that is generically difficult for arbitrary quantum systems, Ryu and Takayanagi [17] beautifully simplified the calculation for field theories that possess classical gravitational duals [15] through their eponymous formula,

$$S(A) = \frac{\text{area}(\tilde{A})}{4G_N}\,. \tag{3.1}$$

In the above, A is a region in the boundary conformal field theory, and $\tilde{A}$ is a minimal-area surface in the bulk gravitational dual such that $\partial A = \partial \tilde{A}$. The Ryu-Takayanagi formula essentially translates the abstract algebra question of taking partial traces of density matrices into a geometric one.

A natural question to ask is exactly how difficult is it to use the Ryu-Takayanagi formula to calculate the entanglement entropy of a boundary region? In arbitrary dimensions, even in the case where A consists of a single simply-connected region, the problem of finding the bulk minimal surface is famously difficult. It is known, for example, that even a discretized version of the problem is **NP**-hard for a bulk that has

three spatial dimensions [58]. The problem simplifies considerably in $\mathrm{AdS}_3/\mathrm{CFT}_2$, where a spacelike slice through the spacetime results in a one-dimensional boundary and a two-dimensional bulk. Simply-connected boundary regions are just intervals that are entirely characterized by their two endpoints, and when the bulk is itself simply-connected, the corresponding bulk minimal-area surface is a single geodesic that is anchored on the boundary at the interval's endpoints. Nevertheless, the story becomes more complicated when A consists of a set of disjoint subregions in the one-dimensional boundary conformal field theory. The question boils down to finding the correct union of geodesics (which run between boundary subregion endpoints) that is altogether both minimal and homologous to A. For A which is the union of n subregions, the brute force solution consists of checking every combination of n geodesics that run between the $2n$ endpoints—a task that scales exponentially in n. Is there a more efficient way to identify the correct union of geodesics, or is the combinatorics of boundary subregions a source of hardness even in a one-dimensional boundary?

We find that there is a strong simplification to a polynomial time algorithm in three-dimensional gravity. Following previous inspiring and precise statements of a connection between the Ryu-Takayanagi conjecture and max-flow/min-cut [25, 59], we first devise a constructive algorithm that reduces the problem of determining the minimal-area surface in the context of $\mathrm{AdS}_3/\mathrm{CFT}_2$ to solving max-flow/min-cut on a graph. Crucially, the latter problem can be solved in polynomial time. We then analyze the computational overhead that is required to reduce the problem to max-flow/min-cut and verify that it requires no more than polynomial time as well.

The organization of the chapter is as follows: In Sec. 3.2 we present the algorithm to identify the bulk minimal surface in mathematical terms, and in Sec. 3.3 we analyze the complexity of this algorithm. In Sec. 3.4, we discuss how our approach generalizes to nontrivial bulk topologies. Finally, in Sec. 3.5 we conclude with a few remarks.

3.2 An algorithm to identify minimal-length bulk surfaces

We begin by precisely stating the problem. Let X be a spatial slice of an asymptotically AdS_3 spacetime, and suppose that X is simply-connected. Further, suppose that X is holographically dual to a CFT state ρ defined on its boundary, ∂X. Let $\{A_i\}_{i=1}^n$ with $n \geq 2$ be a collection of non-empty, simply-connected, closed, disjoint boundary regions, i.e., $A_i \subset \partial X$, $A_i \neq \emptyset$, $A_i = \text{cl}(A_i)$ for $i \in [n]$,[1] and $A_i \cap A_j = \emptyset$ for $i \neq j$. What is the Ryu-Takayanagi surface, i.e., the minimal-length bulk surface $\tilde{A}$ that is homologous to $A = \bigcup_{i=1}^n A_i$? Or, if there are several minimal-length surfaces, what is one of them? Note that we may consider strictly disjoint regions without loss of any generality; two overlapping or touching regions A_i and A_{i+1} may be fused, since the area of any surface that subtends A_i and A_{i+1} separately can only be greater than or equal to the area of a surface that subtends $A_i \cup A_{i+1}$.

Our approach to identifying $\tilde{A}$ is to reformulate the question as a problem on a graph using a construction that is a variation on the one presented in section 3 of [25]. Without loss of generality, suppose that the A_i are numbered according to the order in which they appear going counter-clockwise along ∂X and let $A_i = [a_i, b_i]$, again with respect to counter-clockwise ordering. Next, for each a_i, draw geodesics between it and every b_j so that each geodesic subtends a boundary region $[a_i, b_j]$ whose interior contains zero or an even number of boundary endpoints[2] (Fig. 3.1a). Because X is two-dimensional and simply-connected, these geodesics are precisely the curves that could possibly make up the bulk minimal surface, or in other words, $\tilde{A}$ is a subset of these geodesics. Since each a_i is connected to each of the n endpoints b_j, there are n^2 geodesics in total, and since in the minimal surface each endpoint must be connected to only one other endpoint by a geodesic, $\tilde{A}$ consists of n geodesics. Therefore, the task of finding $\tilde{A}$ amounts to identifying n of the n^2 geodesics whose cumulative length is minimal, subject to the constraint that they must together subtend A.

[1] We use the notation $[n] \equiv \{1, 2, \ldots, n\}$ as in [25].

[2] The graph formed by the endpoints as vertices and the geodesics as edges is an example of a complete bipartite graph.

The set of these geodesics, together with ∂X, partition X into a collection of bulk pieces $\{X_\alpha\}$. Define a weighted graph $\tilde{\Gamma}$ by placing a vertex v_α in each of these pieces, and connect two vertices v_α, $v_{\alpha'}$ with an edge $e_{\alpha\alpha'}$ if the pieces to which they belong share an edge, which is itself a segment of a single geodesic (Fig. 3.1b). Define the weight of $e_{\alpha\alpha'}$ to be the proper length of this geodesic segment, i.e., $\omega(e_{\alpha\alpha'}) = |X_\alpha \cap X_{\alpha'}|$, where $|\cdot|$ denotes proper length in this context. Finally, merge all of the vertices for which $(X_\alpha \cap \partial X) \subset A$ into a single vertex v_A, and similarly merge all of the vertices for which $(X_\alpha \cap \partial X) \subset A^c$, where $A^c = \partial X - A$, into another single vertex v_{A^c} (Fig. 3.1c).

Next, define a cut in the following way[3]:

Definition 3.2.1. *A* k-cut C *is a subset of the edges of a graph* G *such that, upon removal of the edges in* C*,* G *is partitioned into* k *disjoint connected components. The weight of the cut, denoted by* $|C|$*, is defined as the sum of the weights of the edges that constitute the cut, i.e.,*

$$|C| = \sum_{e \in C} \omega(e)\,. \tag{3.2}$$

We then arrive at the following result:

Proposition 3.2.2. *Let* C^* *be a minimal-weight 2-cut that separates* v_A *and* v_{A^c} *in the graph construction above. Then, the union of the geodesic segments to which each* $e_{\alpha\alpha'} \in C^*$ *corresponds is* $\tilde{A}$*, i.e.,*

$$\tilde{A} = \bigcup_{e_{\alpha\alpha'} \in C^*} (X_\alpha \cap X_{\alpha'}). \tag{3.3}$$

Proof: Upon close examination, one can see that the proposition follows from the proof of lemma 3 in [25]. To show this, first recall the definition of the graph from [25]. Their final graph, which we denote by G, is constructed out of boundary-anchored geodesics in the same way as $\tilde{\Gamma}$ (except for the final step where the two sets of boundary vertices are merged); however, the set of geodesics is different.

[3]Note that this definition is slightly different from that of [25].

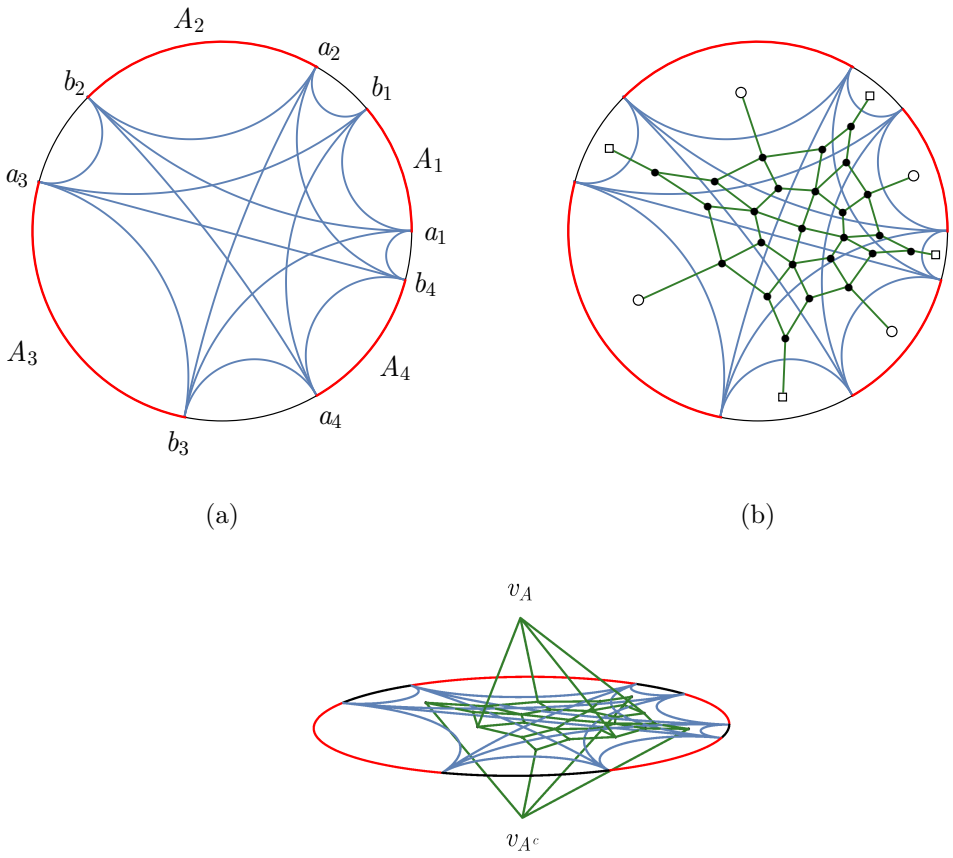

Figure 3.1: Variation on the graph construction from [25] (cf. their figure 4), illustrated for $n = 4$. (a) The boundary regions $A_1, \ldots, A_4$ are shown in red and the geodesics which link each a_i and b_j pair, $1 \leq i, j \leq 4$, are shown in blue. (b) The graph $\tilde{\Gamma}$ is constructed by placing a vertex in each bulk region X_α, and vertices are linked when their respective bulk regions share a geodesic segment as an edge. All of the hollow circular nodes are identified as a single vertex v_A, and all of the hollow square nodes are identified as a single vertex v_{A^c}. This identification is illustrated in (c).

Namely, only those geodesics which constitute the actual minimal surfaces for all possible unions of subsets of $\{A_i\}_{i=1}^n$ are used. Explicitly, for all subsets $I \subseteq [n]$, let $A_I = \bigcup_{i \in I} A_i$ and let $\tilde{A}_I$ be the corresponding Ryu-Takayanagi surface. Then, G is obtained by placing a vertex in each of the pieces into which X is split by $\bigcup_{I \subseteq [n]} \tilde{A}_I$. In particular, note that $\bigcup_{I \subseteq [n]} \tilde{A}_I$ is contained in the set of geodesics that we use to define our graph $\tilde{\Gamma}$.

Next, recall the content of lemma 3 of [25]. For each $I \subseteq [n]$, define the *discrete entropy* as

$$S^*(I) = \min \frac{|C_I|}{4G_N}\,, \tag{3.4}$$

where G_N is Newton's constant and where the minimization is over all k-cuts C_I that separate the $|I|$ boundary vertices corresponding to the pieces X_α for which $(X_\alpha \cap \partial X) \subseteq A_I$ from the rest of the graph. [25, Lemma 3] then states that $S^*(I)$ coincides exactly with the conventional Ryu-Takayanagi entropy $S(I) = |\tilde{A}_I|/4G_N$, and so it also follows that $|\tilde{A}_I| = \min |C_I|$. Moreover, the proof of this lemma further establishes that the minimal cut, call it C_I^*, actually corresponds to the Ryu-Takayanagi surface $\tilde{A}_I$ (or possibly another equivalent surface with the same length if the minimal surface is not unique).

Now, if the two sets of geodesics that are used to define our graph $\tilde{\Gamma}$ and the graph G from [25] had been the same, then the proof would be complete since we would have that $C^* = C^*_{[n]}$. However, we must establish that the minimal cut C^* actually corresponds to a Ryu-Takayanagi surface, given that the set of geodesics that generates $\tilde{\Gamma}$ can be larger than the set that generates G. This result follows from the observation that a minimal cut C_I^* in G correctly identifies a Ryu-Takayanagi surface when $I \neq [n]$. In this case, one may iteratively think of G as being generated by more geodesics than those contained in $\bigcup_{I' \subseteq I} \tilde{A}_{I'}$. This is precisely the case for $\tilde{\Gamma}$, which is itself generated by at least those geodesics which make up $\bigcup_{I \subseteq [n]} \tilde{A}_I$, and so the minimal cut C^* corresponds to $\tilde{A}$.

□

Therefore, in our graph construction, the problem of finding the Ryu-Takayanagi surface for the collection of boundary regions A corresponds to finding C^*, i.e., solving the max-flow/min-cut problem between the two vertices v_A and v_{A^c} on the graph $\tilde{\Gamma}$. Importantly, this is a problem that can be solved efficiently in a time that is polynomial in the number of vertices and edges of $\tilde{\Gamma}$ (see e.g. [60, Chapter 5.4]). Therefore, in order to show that the whole task of finding the Ryu-Takayanagi surface can be completed in polynomial time, all that is left is to establish that the overhead in setting up the graph construction above takes no more than polynomial time in n and that the number of vertices of $\tilde{\Gamma}$ is no more than polynomial in n.

3.3 Complexity analysis

We now revisit the algorithm presented above and verify that that each step has an algorithmic complexity that is polynomial in the number of boundary regions, n. First, we note that the problem can be restated as a decision problem:

Problem 3.3.1. *Given as input*

i. a Riemannian metric $g_{ij}(x)$ together with a coordinate ultraviolet cutoff Λ that describes a simply-connected, asymptotically-hyperbolic, two-dimensional manifold X,

ii. a list of n pairs of points on the conformal boundary of X, $\{[a_i, b_i]\}_{i=1}^n$, that specify n non-empty, disjoint, closed, simply-connected intervals in ∂X, and

iii. a permutation $\sigma : [n] \to [n]$ that identifies n geodesics that connect a_i with $b_{\sigma(i)}$ for $i \in [n]$ and that together subtend the intervals $[a_i, b_i]$,

does there exist another permutation σ' such that

$$\sum_{i=1}^{n} |\gamma_{i,\sigma'(i)}|_\Lambda < \sum_{i=1}^{n} |\gamma_{i,\sigma(i)}|_\Lambda \tag{3.5}$$

up to a numerical precision ϵ, where $\gamma_{i,\sigma'(i)}$ denotes the geodesic between a_i and $b_{\sigma'(i)}$ and $|\gamma_{i,\sigma'(i)}|_\Lambda$ is its proper length with the cutoff Λ in place? (Assume that the a_i and b_i have enough digits of precision to compute at the global precision ϵ.)

Of course, the decision problem can be answered by carrying out the algorithm above to actually find the minimal surface. One reason for writing out this restatement it to clearly identify two sources of algorithmic complexity: complexity in n, the number of boundary regions, as well as the numerical complexity that is a consequence of having to compute real-valued geometric quantities up to precision ϵ. We will focus on the complexity in n, but it should be understood that the overall complexity has some multiplicative scaling $O(f(\epsilon))$ which depends on the numerical techniques that one uses to compute geometric quantities.

A dual for a dual

As a preliminary step, it is useful to define a second graph, Γ, by placing a vertex at every point where two or more geodesics intersect and at each of the a_i and b_i. Connect two vertices with an edge if they are adjacent to each other on a single geodesic, and also add an edge in between each adjacent boundary endpoint (so that the vertices at a_1 and b_1, at b_1 and a_2, at a_2 and b_2, etc. gain an additional edge connecting them). With this definition, $\tilde{\Gamma}$ is (up to the merger of the boundary vertices into v_A and v_{A^c}) the dual graph of Γ, which will be useful for counting (Fig. 3.2).

Finding the geodesics

Since X is two-dimensional, any boundary-anchored geodesic can be parametrized by two real numbers, for instance, its two endpoints on ∂X. As such, drawing the n^2 geodesics between the a_i and the b_j consists of solving the geodesic equation in terms of the two free parameters and then listing the n^2 specific solutions. When the geodesic equation has a closed-form solution, obtaining each geodesic is an $O(1)$ overhead. For example, when X is the hyperbolic plane, it is straightforward to show that geodesics in the Poincaré disk are circular arcs that are normal to the boundary, and a specific arc can be labelled by its endpoints. Here, one must only solve the geodesic equation once with its endpoints as free parameters. If the geodesic equation does not have a closed-form solution and/or one works numerically, then

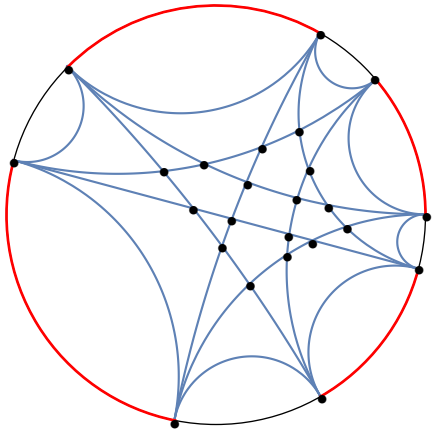

Figure 3.2: The graph Γ is constructed by placing a vertex at each point where two or more geodesics intersect and at the boundary interval endpoints a_i and b_j. Vertices are connected by the geodesic and boundary segments on which they lie.

constructing and digitally representing each geodesic will have some complexity that depends on ϵ. Note, however, that this is independent of the number of boundary regions, n. Therefore, the scaling of this step is $O(n^2)$. We will suppose that the output of this subroutine is a list of functions $\gamma_{i,j} : [0,1] \to X$ that parametrize the geodesics, i.e., $\gamma_{i,j}(0) = a_i$, $\gamma_{i,j}(1) = b_j$, and whose runtime is independent of n.

Number of vertices in Γ

Let V be the number of vertices in Γ. This is equal to $2n$ (the number of boundary region endpoints) plus the number of times that the n^2 geodesics intersect each other. We can upper bound the latter quantity by $p \cdot \binom{n^2}{2}$, which would be the number of intersections if every geodesic intersected every other geodesic at most p times. We must assume that p is a bounded constant. Generically, we expect that $p = 1$; in the case where X is simply-connected and has nonpositive curvature everywhere, then $p = 1$ is implied by the Cartan-Hadamard Theorem, which guarantees that pairs of points are connected by a unique geodesic (see, for instance, [61, Theorem 4.5]).

This point is further discussed in Sec. 3.6. With this assumption on p, we have that

$$V \leq 2n + p\binom{n^2}{2} = 2n + \tfrac{1}{2}pn^2(n^2-1) = O(n^4)\,. \tag{3.6}$$

Number of edges in Γ *and* $\tilde{\Gamma}$

Let E be the number of edges in Γ. This is equal to $2n$ (the number of edges that lie on ∂X) plus the number of geodesic segments in the interior of X. Note that since $\tilde{\Gamma}$ is, up to the vertex mergers, the dual graph of $\tilde{\Gamma}$, this latter quantity is also the number of edges in $\tilde{\Gamma}$, which we denote by $\tilde{E}$.

Similarly to the counting of vertices above, the largest number of interior edges is upper bounded by the number of interior edges in a configuration where each geodesic is intersected p times by the n^2-1 other geodesics in distinct locations. In this case, each geodesic is divided into $p(n^2-1)+1$ segments, and so

$$\begin{aligned} \tilde{E} &\leq n^2\left[p(n^2-1)+1\right] = O(n^4)\,, \\ E &\leq 2n + n^2\left[p(n^2-1)+1\right] = O(n^4)\,. \end{aligned} \tag{3.7}$$

Number of vertices in $\tilde{\Gamma}$

The number of vertices in $\tilde{\Gamma}$ is the number of faces in Γ (or equivalently the number of pieces X_α), which we denote by F, less $2(n-1)$ to account for the vertices that are merged into v_A and v_{A^c}. Since Γ is a planar graph, we can use its Euler characteristic to bound F. From $V - E + F = 2$, it follows that

$$F = 2 + E - V \leq 2 + E = n^2\left[p(n^2-1)+1\right] + 2n + 2 = O(n^4)\,. \tag{3.8}$$

As such, the number of vertices and edges in $\tilde{\Gamma}$ altogether scales like $O(n^4)$.

Connectivity of the vertices and edge weights

So far we have established that the size of Γ and $\tilde{\Gamma}$ scales like $O(n^4)$, but we must also establish that the graphs can be constructed in a number of steps that is polynomial in n. In other words, we must be able to locate vertices, determine their connectivity, and compute edge weights efficiently.

Roughly, locating vertices in Γ amounts to checking if each pair of geodesics intersects, where each check is a constant overhead if closed-form solutions for the geodesics are known, or some ϵ-dependent overhead if one works numerically. This task scales like $\binom{n^2}{2} = O(n^4)$. Then, as noted above, the faces of Γ are the vertices of $\tilde{\Gamma}$. The weight of an edge in Γ (and also $\tilde{\Gamma}$) is given by the proper length of its corresponding geodesic segment, and so computing this weight amounts to performing a line integral along the geodesic segment. At worst, if the exact antiderivative is unknown, evaluating this integral numerically up to a fixed numerical accuracy is again a computational task that must be performed less than $\tilde{E}$ times, and so the algorithmic complexity of this step scales like $O(n^4)$.

To be a bit more concrete, let us sketch an algorithm to construct a digital representation of Γ. Represent Γ with a $V \times V$ upper-triangular matrix M, and denote the vertices of Γ by w_α. For $\alpha < \beta$, the entries of M will be $M_{\alpha\beta} = -1$ if w_α and w_β are adjacent vertices on the boundary, $M_{\alpha\beta} = \omega(e_{\alpha\beta})$ if w_α and w_β are connected via a shared geodesic segment, and zero otherwise. Let the first $2n$ vertices be the boundary vertices, i.e., $w_{2\alpha-1} \equiv a_\alpha$ and $w_{2\alpha} \equiv b_\alpha$ for $1 \leq \alpha \leq n$. For each geodesic $\gamma_{i,j}$, $1 \leq i, j \leq n$, we will construct a list $L_{i,j}$ whose entries are pairs (w_α, s_α) which identify the vertices w_α that lie on $\gamma_{i,j}$, as well as the $s_\alpha \in [0, 1]$ which specifies the location $\gamma_{i,j}(s_\alpha)$ in X (and hence also on the geodesic itself) where the vertex lies. Each $L_{i,j}$ can therefore be initialized with two elements,

$$L_{i,j} = \{(w_i, 0), (w_{j+1}, 1)\} \ . \tag{3.9}$$

The following pseudo-code then sketches how to construct the $L_{i,j}$ and M. A bold index will denote a composite index, i.e., $\mathbf{i} \equiv i, j$.

```
% First fill in M for boundary vertices
for λ from 1 to 2n−1
    M_{λ λ+1} = −1
end for
M_{1 2n} = −1
```

% Next fill in the interior vertices

$\kappa = 2n + 1$

for **i** from 1 to n^2 %*

  % Build up the $L_{\mathbf{i}}$

  for $\mathbf{j} > \mathbf{i}$

    if $\gamma_{\mathbf{i}}$ and $\gamma_{\mathbf{j}}$ intersect

      label this vertex w_κ

      find the intersection location $s_\kappa^{(\mathbf{i})}$ on $\gamma_{\mathbf{i}}$ and $s_\kappa^{(\mathbf{j})}$ on $\gamma_{\mathbf{j}}$ %**

      append $(w_\kappa, s_\kappa^{(\mathbf{i})})$ to $L_{\mathbf{i}}$ and $(w_\kappa, s_\kappa^{(\mathbf{j})})$ to $L_{\mathbf{j}}$;

      κ++

    end if

  end for

  sort $L_{\mathbf{i}}$ according to increasing s_α

  % Compute edge weights and fill in M

  for k from 1 to (Length[$L_{\mathbf{i}}$]-1)

    let $(w_{\alpha_k}, s_{\alpha_k})$ be the k^{th} element of $L_{\mathbf{i}}$

    compute the length of $\gamma_{\mathbf{i}}$ from s_{α_k} to $s_{\alpha_{k+1}}$, i.e., $\omega(e_{\alpha_k\alpha_{k+1}})$ %**

    $M_{\alpha_k\alpha_{k+1}} = \omega(e_{\alpha_k\alpha_{k+1}})$

  end for

end for

Note that the nested loop beginning on the line with an asterisk (*) executes $O(n^4)$ times as expected. The steps that may contribute numerical ϵ-dependent overhead occur the lines with two asterisks (**). Finally, the various array accesses and other tasks (such as the sorting operation) will only contribute a polynomial number of steps.

We therefore ultimately find that the time it takes to set up our graph construction scales like $O(n^4)$, and that $\tilde{\Gamma}$ itself has a number of vertices and a number of edges that are each $O(n^4)$. The complexity of max-flow/min-cut is $O(\tilde{E}\tilde{V}^2)$ [60], and so the

overall complexity of our algorithm is $O(n^{12})$. As such, the algorithmic complexity of finding the Ryu-Takayanagi surface for n boundary regions is poly(n) as claimed.

3.4 Other bulk topologies

The algorithm as described above applies to simply-connected bulk geometries. In situations where the bulk is topologically nontrivial, there is a new parameter which the algorithm could scale with, namely, the genus q of the bulk topology. This is because it is no longer true that the minimal surface that is homologous to a single simply-connected boundary region is necessarily made up of a single geodesic. Consequently, the number of geodesics changes from n^2 to some $O(f(q)n^2)$. But, once all of the possible minimal surfaces are determined, one can simply continue apace with the max-flow/min-cut algorithm as before.

The scaling with q will not change the scaling with the number of boundary intervals; it enters as another independent multiplicative factor. As a last discussion item, let us estimate what the worst scaling with q could be. First, consider drawing an extremal path $\mathcal{C}$ between two boundary endpoints a_i and b_j when there are q punctures in X (Fig. 3.3). In theory, provided that the geodesics exist, we could choose to include anywhere from zero up to all q of the punctures in $\text{int}([a_i, b_j] \cup \mathcal{C})$, where $\text{int}(\cdot)$ denotes the interior of a closed curve. Then, noting that there may be up to $\binom{q}{k}$ ways to include k punctures, we identify up to $\sum_{k=0}^{q} \binom{q}{k} = 2^q$ geodesics in this way. However, each time that a puncture is included in $\text{int}([a_i, b_j] \cup \mathcal{C})$, we also must draw a geodesic around the puncture so that the total (multiply-connected) extremal curve is homologous to $[a_i, b_j]$. So, we must also consider the set of all geodesics that enclose anywhere from one to all q of the punctures, where there are $\binom{q}{k'}$ ways to enclose k' punctures. This gives us another $\sum_{k'=1}^{q} \binom{q}{k} = 2^q - 1$ geodesics. Repeating this analysis for every pair of points a_i and b_j, we conclude that there are at most $2^q n^2$ geodesics that connect boundary endpoints and $2^q - 1$ geodesics that enclose punctures (which remain the same for every pair of boundary endpoints). Therefore, there will be at most $2^q n^2 + 2^q - 1$ geodesics that seed the rest of the

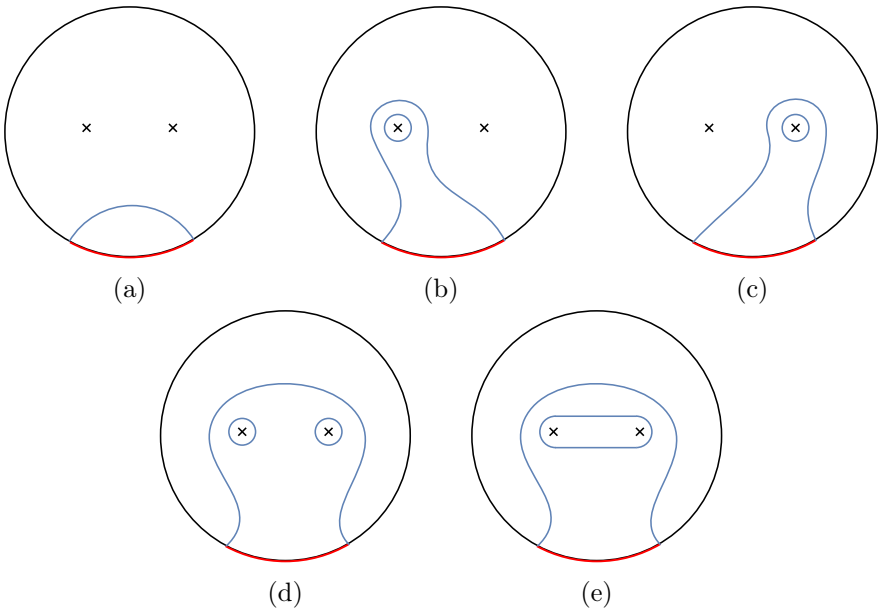

Figure 3.3: Possible ways that the extremal surface for a single simply-connected boundary region could include or exclude punctures, illustrated for $q = 2$. The boundary interval is shown in red, the extremal curve is shown in blue, and punctures are represented by crosses.

algorithm. As expected, the scaling in n is unchanged, but the scaling in q can be very large indeed.

3.5 Conclusion

We have demonstrated that the task of finding the Ryu-Takayanagi surface in three-dimensional gravity for a collection of n boundary subregions has polynomial complexity by exhibiting an algorithm that completes this task in poly(n) time. The algorithm consists of converting the geometric problem into a graph-theoretic problem whose solution is given by the max-flow/min-cut between two vertices on a graph. The procedure is essentially a discretization of Freedman and Headrick's bit thread model [59] with the flow being between the boundary subregions $\bigcup_{i=1}^{n} A_i \equiv A$ and the rest of the boundary, $\partial X - A$.

It should be noted that performing this calculation holographically in many ways complements the computation of entanglement entropy using only CFT_2 techniques, e.g. [62,63]. On one hand, results for simply-connected boundary regions such as the

Cardy-Calabrese formula $S = \frac{c}{3} \log \frac{l}{\Lambda}$ do not extend to multiply-disjoint regions. On the other hand, entanglement entropies of simply-connected regions are the "primitives" in our algorithm, and so in pure AdS_3/CFT_2 for example, one can use the Cardy-Calabrese formula to avoid finding geodesics altogether. By extension, if you knew how to compute S for simply-connected regions in the boundary for some given holographic CFT_2 state, then assuming the Ryu-Takayanagi conjecture, computing geodesics becomes unnecessary since geodesic length is automatically given by $4GS$. See, for example, [64], which demonstrates powerful algebraic methods to compute these geodesic lengths in a broad class of CFT states.

Some future interesting directions would be to use our line of reasoning to clearly delineate the sources of complexity that make the higher dimensional case **NP**-Hard. For example, it is plausible that the combinatorial aspect of the problem is in general not difficult, but rather that the difficulty arises from the fact that simply-connected boundary regions do not have a canonical shape in higher dimensions. In a related way, it would also be interesting to extend the algorithm to higher dimensions, but where boundary subregions are restricted to only have certain shapes, e.g., filled S^{d-1} spheres on a S^d conformal boundary. This sort of setting is important for holographic derivations of the Einstein equations [65–67] among other applications.

It has also been pointed out that the relationship of holographic entanglement entropy to max-flow/min-cut may extend to covariant formulations [32, 59, 68, 69]. If this is fully established, it would certainly be interesting to see whether our analyses can be extended past minimal surfaces to maximin formulations to arrive at a similar style of conclusion.

Acknowledegements

We thank Adam Bouland, Wilson Brenna, Matthew Headrick, John Preskill, and Michael Walter for helpful discussions and suggestions. This material is based upon work supported in part by the following funding sources: N.B. was supported in part by the DuBridge Postdoctoral Fellowship, by the Institute for Quantum Information

and Matter, a NSF Physics Frontiers Center (NFS Grant PHY-1125565) with support of the Gordon and Betty Moore Foundation (GBMF-12500028). A.C.-D. was supported by the NSERC Postgraduate Scholarship program and by the Gordon and Betty Moore Foundation through Grant 776 to the Caltech Moore Center for Theoretical Cosmology and Physics. This work was supported by the U.S. Department of Energy, Office of Science, Office of High Energy Physics, under Award Number DE-SC0011632.

3.6 Intersections of geodesics

Here we consider the question of how many times two distinct boundary-anchored geodesics in X can intersect. When X has nonpositive curvature everywhere and is simply-connected, such as the case where X is the hyperbolic plane, then the Cartan-Hadamard theorem implies that every pair of points is connected by a unique geodesic. We can use this fact to obtain the following result:

Proposition 3.6.1. *Let X by a simply-connected Riemannian manifold with non-positive curvature everywhere. Then, two distinct geodesics can intersect each other at most once.*

Proof: We establish the proof by contradiction. Let $\mathcal{C}_1$ and $\mathcal{C}_2$ be two distinct geodesics, and suppose that they intersect more than once. Let p_1 and p_2 be two intersection points, and denote the segment of $\mathcal{C}_1$ (resp. $\mathcal{C}_2$) that connects p_1 and p_2 by $\mathcal{S}_1$ (resp. $\mathcal{S}_2$). The lengths of $\mathcal{S}_1$ and $\mathcal{S}_2$ cannot be the same. This is because the Cartan-Hadamard theorem holds, and so the geodesic that connects p_1 and p_2 is unique. Without loss of generality, suppose that $|\mathcal{S}_1| < |\mathcal{S}_2|$. But then, $(\mathcal{C}_2 - \mathcal{S}_2) \cup \mathcal{S}_1$ is shorter than $\mathcal{C}_2$, which contradicts the assumption that $\mathcal{C}_2$ is a geodesic.

□

Note that this proposition does not exclude the case where $\mathcal{C}_1$ and $\mathcal{C}_2$ overlap on a finite interval. However, such behaviour does not change the scaling of the number of vertices in Γ if we only place vertices at the points where the geodesics first

meet. Also note that the result holds for Riemannian manifolds X of any dimension, and X can be relaxed to a metric space if the curvature is taken to be Alexandrov curvature [61].

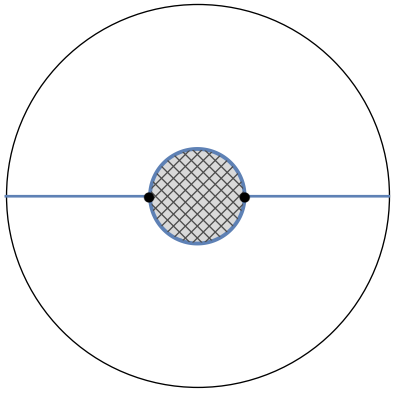

Figure 3.4: Spatial slice through the (1+2)-dimensional BTZ spacetime. Two antipodal points (shown in black) on the black hole (the hatched disk) are connected by two distinct geodesics (shown in blue).

Cases where two boundary-anchored geodesics can intersect more than once are necessarily cases where there exist points in X such that the geodesic connecting them is not unique. An example of such a configuration occurs when X is a slice of the BTZ black hole spacetime [70] (and is therefore not simply-connected). Two points in the bulk that are antipodal with respect to the black hole are connected by geodesics of the same length that wrap around either side of the black hole (Fig. 3.4). The boundary-anchored geodesics on which the two points lie share their boundary endpoints, however, and so this particular configuration is excluded from the configurations that we consider. As such, it seems reasonable to expect that cases where two boundary-anchored geodesics can intersect more than once and which are allowed by the problem under consideration, if they exist, are pathological and could be excluded with an appropriate generic condition.

Chapter 4

CONSISTENCY CONDITIONS FOR AN ADS/MERA CORRESPONDENCE

The Multi-scale Entanglement Renormalization Ansatz (MERA) is a tensor network that provides an efficient way of variationally estimating the ground state of a critical quantum system. The network geometry resembles a discretization of spatial slices of an AdS spacetime and "geodesics" in the MERA reproduce the Ryu–Takayanagi formula for the entanglement entropy of a boundary region in terms of bulk properties. It has therefore been suggested that there could be an AdS/MERA correspondence, relating states in the Hilbert space of the boundary quantum system to ones defined on the bulk lattice. Here we investigate this proposal and derive necessary conditions for it to apply, using geometric features and entropy inequalities that we expect to hold in the bulk. We show that, perhaps unsurprisingly, the MERA lattice can only describe physics on length scales larger than the AdS radius. Further, using the covariant entropy bound in the bulk, we show that there are no conventional MERA parameters that completely reproduce bulk physics even on super-AdS scales. We suggest modifications or generalizations of this kind of tensor network that may be able to provide a more robust correspondence.

This chapter was published as Ref. [3], N. Bao, C. Cao, S. M. Carroll, A. Chatwin-Davies, N. Hunter-Jones, J. Pollack, and G. N. Remmen, "Consistency conditions for an AdS multiscale entanglement renormalization ansatz correspondence," Phys. Rev. D ***91*** *(2015) 125036,* `arXiv:1504.06632`.

4.1 Introduction

The idea that spacetime might emerge from more fundamental degrees of freedom has long fascinated physicists. The holographic principle suggests that a $(D+1)$-dimensional spacetime might emerge from degrees of freedom in a D-dimensional the-

ory without gravity [71, 72]. While a completely general implementation of this idea is still lacking, the AdS/CFT correspondence provides a specific example in which to probe the holographic emergence of spacetime. AdS/CFT is a conjectured correspondence between D-dimensional conformal field theories (CFTs) in Minkowski space and $(D+1)$-dimensional asymptotically anti-de Sitter (AdS) spacetimes [15, 16, 73]. An intriguing aspect of this duality is the Ryu–Takayanagi formula [17, 74], according to which the entanglement entropy of a region B on the boundary is proportional to the area of a codimension-two extremal surface $\tilde{B}$ embedded in the bulk curved spacetime whose boundary is B:

$$S(B) = \frac{\text{area}(\tilde{B})}{4G} + \text{corrections}. \tag{4.1}$$

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

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

Meanwhile, from a very different perspective, tensor networks have arisen as a useful way to calculate quantum states in strongly-interacting many-body systems [36]. One significant example is the Multi-scale Entanglement Renormalization Ansatz (MERA) [75], which is relevant for critical (gapless) systems, i.e., CFTs. Starting from a simple state in a low-dimensional Hilbert space, acting repeatedly with fixed tensors living on a network lattice produces an entangled wave function for the quantum system of interest; varying with respect to the tensor parameters efficiently computes the system's ground state.

Working "backwards" in the MERA, starting with the ground state and gradually re-

moving entanglement, produces a set of consecutively renormalized quantum states. This process reveals a renormalization direction along the graph, which may be thought of as an emergent radial direction of space. As pointed out by Swingle [76], the MERA graph can serve as a lattice discretization of spatial slices of AdS. Furthermore, one can use the MERA to calculate the entanglement entropy of regions of the original (boundary) critical system; this calculation amounts to tracing over bonds in the tensor network that cross the causal cone of the boundary region. The causal cone is a sort of extremal surface for the MERA, motivating comparison to the Ryu–Takayanagi formula.

It is therefore natural to conjecture that the MERA provides a concrete implementation of the emergence of spacetime, in the form of a correspondence between boundary and bulk regions reminiscent of AdS/CFT [76]. Such an AdS/MERA correspondence would be extremely useful, since the basic building blocks of the MERA are discrete quantum degrees of freedom from which quantities of physical interest may be directly calculated. Some specific ideas along these lines have recently been investigated [31, 77–79].

In this paper, we take a step back and investigate what it would mean for such a correspondence to exist and the constraints it must satisfy in order to recover properties we expect of physics in a bulk emergent spacetime. After reviewing the MERA itself and possible construals of the AdS/MERA correspondence in the next section, in Sec. 4.3 we then derive relationships between the MERA lattice and the geometry of AdS. We find that the MERA is unable to describe physics on scales shorter than the AdS radius. In Sec. 4.4 we explore constraints from calculating the entanglement entropy of regions on the boundary, in which we are able to relate MERA parameters to the central charge of the CFT. Finally, in Sec. 4.5 we apply the covariant entropy (Bousso) bound to regions of the bulk lattice. In the most naïve version of the AdS/MERA correspondence, we find that no combination of parameters is consistent with this bound, but we suggest that generalizations of the tensor network may be able to provide a useful correspondence.

4.2 AdS/MERA

Let us begin by recalling the definition and construction of the MERA. We will then introduce the AdS/MERA correspondence and discuss the motivation for and consequences of this proposal.

4.2.1 Review of the MERA

The MERA is a particular type of tensor network that provides a computationally efficient way of finding the ground states of critical quantum many-body systems, i.e. CFTs, in D dimensions. (For a recent review of tensor networks in general, see Ref. [36]. Detailed analyses of the MERA are given in [75, 80, 81] and references therein.) In this work, we restrict our attention to the case $D = 1 + 1$.

The MERA tensor network is shown in Fig. 4.1. The quantum system being modeled by the MERA lives at the bottom of the diagram, henceforth "the boundary" in anticipation of the AdS/MERA connection to be explored later. We can think of the tensor network as a quantum circuit that either runs from the top down, starting with a simple input state and constructing the boundary state, or from the bottom up, renormalizing a boundary state via coarse-graining. One defining parameter of the MERA is the rescaling factor k, defining the number of sites in a block to be coarse-grained; in Fig. 4.1 we have portrayed the case $k = 2$. The squares and triangles are the tensors: multilinear maps between direct products of vector spaces. Each line represents an index i of the corresponding tensor, ranging over values from 1 to the "bond dimension" χ. The boundary Hilbert space $\mathcal{H}_{\text{boundary}} = V^{\otimes \mathcal{N}_{\text{boundary}}}$ is given by a tensor product of $\mathcal{N}_{\text{boundary}}$ individual spaces V, each of dimension χ. (In principle the dimension of the factors in the boundary could be different from the bond dimension of the MERA, and indeed the bond dimensions could vary over the different tensors. We will assume these are all equal.)

As its name promises, the MERA serves to renormalize the initial boundary state via coarse-graining. If we were to implement the MERA for only a few levels, we would end up with a quantum state in a smaller Hilbert space (defined on a fixed

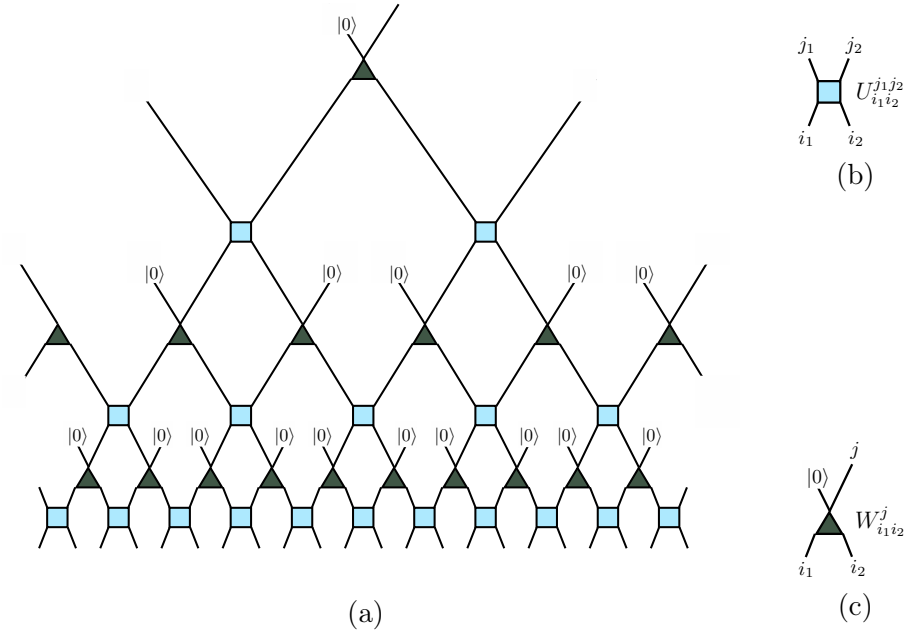

Figure 4.1: (a) Basic construction of a $k = 2$ MERA (2 sites renormalized to 1). (b) The squares represent disentanglers: unitary maps that, from the moving-upward perspective, remove entanglement between two adjacent sites. (c) The triangles represent isometries: linear maps that, again from the moving-upward perspective, coarse-grain two sites into one. Moving downward, we may think of isometries as unitary operators that, in the MERA, map a state in $V \otimes |0\rangle$ into $V \otimes V$. The i and j labels in (b) and (c) represent the tensor indices of the disentangler and isometry.

level of the tensor network), retaining some features of the original state but with some of the entanglement removed. However, we can also run the MERA backwards, to obtain a boundary state from a simple initial input. By varying the parameters in the individual tensors, we can look for an approximation of the ground state of the CFT on the boundary. Numerical evidence indicates that this process provides a computationally efficient method of constructing such ground states [81, 82].

The tensors, or gates, of the MERA come in two types. The first type are the disentanglers, represented by squares in Fig. 4.1. These are unitary maps $U : V \otimes V \to V \otimes V$, as in Fig. 4.1 (b). The name comes from thinking of moving upward

through the network, in the direction of coarse-graining, where the disentanglers serve to remove local entanglement; as we move downward, of course, they take product states and entangle them. The second type of tensors are the isometries, represented by triangles. From the moving-downward perspective these are linear maps $W : V \to V \otimes V$; moving upward, they implement the coarse-graining, see Fig. 4.1 (c). The isometries are subject to the further requirement that $W^\dagger W = I_V$, where I_V is the identity map on V, and $WW^\dagger = P_A$, where P_A is a projector onto some subspace $A \subset V \otimes V$. From the top-down perspective, we can also think of the isometries as bijective unitary operators $W_U : V \otimes V \to V \otimes V$, for which a fixed "ancilla" state (typically the ground state $|0\rangle$) is inserted in one of the input factors, as shown in Fig. 4.1 (c). More generally, isometries could map $q < k$ sites onto k sites, $W : V^{\otimes q} \to V^{\otimes k}$.

The MERA is not the simplest tensor network which implements coarse-graining. For instance, the tree tensor network [83] (also considered in a holographic context in Ref. [77]), similar to MERA but without any disentanglers, also implements coarse-graining. However, tensor networks without disentanglers fail to capture the physics of systems without exponentially-decaying correlations, and consequently cannot reproduce a CFT ground state.

An example that invites analysis with a MERA is the transverse-field Ising model [84]. In $1+1$ dimensions, the model describes a chain of spins with nearest-neighbor interactions subject to a transverse magnetic field. Its Hamiltonian is

$$\hat{H} = -J \sum_i \hat{\sigma}_i^z \hat{\sigma}_{i+1}^z - h \sum_i \hat{\sigma}_i^x \,, \tag{4.2}$$

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

true real-space renormalization, producing a lattice of spins that is less dense than the preceding lattice by a factor of q/k.

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

For geometric considerations, it is useful to abstract away all of the information about unitary operators and to draw a MERA as a graph as shown in Fig. 4.2. In such a graph, we only indicate the connectivity of sites at any given level of coarse-graining as well as the connectivity of sites under renormalization group flow.

4.2.2 An AdS/MERA Correspondence?

The possibility of a correspondence between AdS and the MERA was first proposed by Swingle in Ref. [76], where it was noted that the MERA seems to capture certain key geometric features of AdS. At the most basic level, when viewed as a graph with legs of fixed length, a MERA may be thought of as a discretization of the hyperbolic plane, which is a spatial slice of AdS_3. In this discretization, the base of the MERA tree lies on the boundary of the AdS slice and the MERA lattice sites fill out the bulk of the slice [76, 85].

Interestingly, the structure of a MERA is such that it seems to go beyond a simple discretization of the hyperbolic plane. Certain discrete paths in the MERA naturally reproduce geodesics of the hyperbolic plane [76, 86]. Moreover, this phenomenon makes it possible to understand the computation of CFT entanglement entropy using

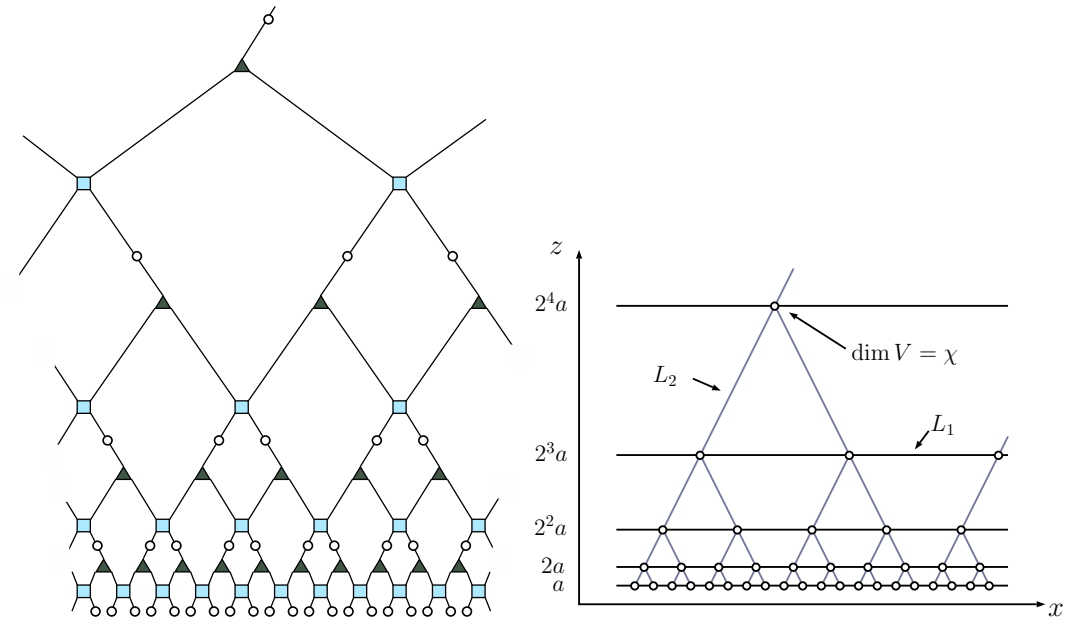

Figure 4.2: (a) A $k = 2$ MERA, and (b) the same MERA with its disentanglers and isometries suppressed. The horizontal lines in the graph on the right indicate lattice connectivity at different renormalization depths, and the vertical lines indicate which sites at different depths are related via coarse-graining due to the isometries. Each site, represented by a circle, is associated with a Hilbert space V with bond dimension χ. In the simplest case, a copy of the same Hilbert space is located at each site. When assigning a metric to the graph on the right, translation and scale invariance dictate that there are only two possible length scales: a horizontal proper length L_1 and a vertical proper length L_2.

a MERA as a discrete realization of the Ryu–Takayanagi formula [87]. These and other examples [76,86] seem to suggest that a MERA may in fact be elucidating the structural relationship between physics on the boundary of AdS and its bulk.

In this work we take the term "AdS/MERA correspondence" to mean more than simply a matching of graph geometry and continuous geometry. In the spirit of the AdS/CFT correspondence, we suppose that (at least some aspects of) both boundary and bulk physics are described by appropriate Hilbert spaces $\mathcal{H}_{\text{boundary}}$ and $\mathcal{H}_{\text{bulk}}$ respectively, which must have equal dimensions. A full AdS/MERA correspondence would then be a specification of these Hilbert spaces, as well as a prescription which makes use of the MERA to holographically map states and operators in $\mathcal{H}_{\text{boundary}}$ to corresponding states and operators in $\mathcal{H}_{\text{bulk}}$ and vice-versa. To preserve locality

in the bulk and the symmetries of AdS, it is natural to identify $\mathcal{H}_{\text{bulk}}$ with the tensor product of individual spaces V_{bulk}, each located at one site of the MERA. If it exists, this correspondence provides a formulation of bulk calculations in terms of the MERA. An AdS/MERA correspondence should allow us to, for example, calculate bulk correlation functions, or bulk entanglement entropies using tools from or the structure of the MERA.

There is one straightforward way to construct such a map $\mathcal{H}_{\text{boundary}} \leftrightarrow \mathcal{H}_{\text{bulk}}$. We have noted that the isometries $W : V \to V \otimes V$ can be thought of as unitaries $W_U : V \otimes V \to V \otimes V$ by imagining that a fixed ancillary state $|0\rangle$ is inserted in the first factor; for a k-to-one MERA, one would insert $k-1$ copies of the $|0\rangle$ ancilla at each site to unitarize the isometries. From that perspective, running upwards in the tensor network provides a map from the MERA ground state on the boundary to a state $|0\rangle^{\otimes(k-1)\mathcal{N}_{\text{bulk}}} \in V^{\otimes(k-1)\mathcal{N}_{\text{bulk}}}$, where at each isometry there is a copy of $V^{\otimes(k-1)}$ and $\mathcal{N}_{\text{bulk}}$ denotes the number of bulk lattice sites, excluding the boundary layer. As we ultimately show in Sec. 4.5, one has $\mathcal{N}_{\text{boundary}} = (k-1)\mathcal{N}_{\text{bulk}}$. We can then identify $\mathcal{H}_{\text{boundary}} = \mathcal{H}_{\text{bulk}} = V^{\otimes\mathcal{N}_{\text{boundary}}}$ and think of the tensor network as a quantum circuit providing a map between arbitrary states $\mathcal{H}_{\text{boundary}} \to \mathcal{H}_{\text{bulk}}$. In this construction, the MERA ground state on the boundary gets mapped to the factorized bulk state $|0\rangle^{\otimes(k-1)\mathcal{N}_{\text{bulk}}}$, but other boundary states will in general produce entangled states in the bulk (keeping the tensors themselves fixed).

Something very much like this construction was proposed by Qi [77], under the name "Exact Holographic Mapping" (EHM). That work examined a tensor network that was not quite a MERA, as no disentanglers were included, only isometries. As a result, while there is a map $\mathcal{H}_{\text{boundary}} \to \mathcal{H}_{\text{bulk}}$, the boundary state constructed by the tensor network does not have the entanglement structure of a CFT ground state. In particular, it does not seem to reproduce the Ryu–Takayanagi formula in a robust way. Alternatively, we can depart from Qi by keeping a true MERA with the disentanglers left in, in which case the bulk state constructed by the quantum circuit has no entanglement: it is a completely factorized product of the ancilla states. Such

a state doesn't precisely match our expectation for what a bulk ground state should look like, since there should be at least some entanglement between nearby regions of space.

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

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

4.3 MERA and Geometry

If a MERA is a truly geometrical object that describes a slice of AdS, then the graph geometry of a MERA should give the same answers to geometric questions as the continuous geometry of a slice of AdS. Here, we reconsider the observation by Swingle [76, 86] that certain trajectories on the MERA coincide with trajectories in AdS and we investigate the constraints that this correspondence places on the graph metric of the MERA. We find that a MERA necessarily describes geometry on super-AdS length scales, moreover, there is no redefinition of the MERA coordinates that results in the proper distance between MERA sites mapping to any sub-AdS length scale.

4.3.1 Consistency conditions from matching trajectories

In order to speak of graph geometry, one must put a metric on the MERA graph, i.e., one must assign a proper length to each bond in the graph of Fig. 4.2. Presumably, the metric should originate from correlations between the sites in the MERA. In the absence of an explicit identification of the origin of the graph metric, however, at least in the case of a MERA describing the ground state of a CFT, it is sensible to identify two length scales. Explicitly, we must assign a proper length L_1 to horizontal bonds and a proper length L_2 to vertical bonds. Indeed, translational and conformal invariance guarantee that these are the only two length scales in any graph metric one can assign to a MERA for which an AdS/MERA correspondence exists. In particular, the ground state of a CFT is translation invariant, so each horizontal bond in the finest (UV-most) lattice should have the same proper length so as to respect this symmetry. Self-similarity at all scales then requires that any horizontal bond at any level of renormalization have this same proper length. There is no *a priori* reason why the vertical bonds should share the proper length of the horizontal bonds and indeed we will see that their proper length will be different. However, again by self-similarity and translation invariance, all vertical bonds must be assigned the same proper length.

The observation in Ref. [76] that certain paths in the MERA graph coincide with corresponding paths in slices of AdS is what established the possibility of an AdS/MERA correspondence. Here we will carefully examine these paths and determine what constraints the requirements that they match place on MERA parameters, i.e., on the bond lengths L_1 and L_2 and on the rescaling factor k.

Consider a constant-time slice of AdS_3 with the following metric:

$$\mathrm{d}s^2 = \frac{L^2}{z^2}(\mathrm{d}z^2 + \mathrm{d}x^2). \tag{4.3}$$

We will compare the proper lengths of straight horizontal lines and geodesics in the AdS slice to the proper lengths of the corresponding paths in the MERA graph. In the AdS slice, let γ_1 be a straight horizontal line ($\mathrm{d}z = 0$) sitting at $z = z_0$ with

coordinate length x_0. Let γ_2 be a geodesic whose endpoints lie near the boundary $z = 0$ and are separated by a coordinate distance x_0 at the boundary. In this choice of coordinates, such a geodesic looks like a semicircle (see Fig. 4.3). It is a straightforward computation to show that the proper lengths of these curves are

$$|\gamma_1|_{\text{AdS}} = \frac{L}{z_0} x_0 \qquad \text{and} \qquad |\gamma_2|_{\text{AdS}} = 2L \ln\left(\frac{x_0}{a}\right) . \tag{4.4}$$

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

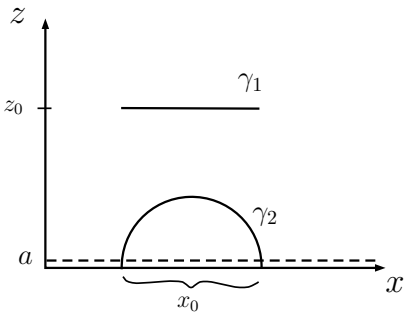

Figure 4.3: A horizontal line (γ_1) and a geodesic (γ_2) in a spatial slice of AdS_3.

We fix L_1 and L_2 by comparing γ_1 and γ_2 to horizontal lines and "geodesics" in the MERA, respectively. Consider two sites in a horizontal lattice at depth m (i.e., m renormalizations of the UV-most lattice) and separated by a coordinate distance x_0 in the coordinate system shown in Fig. 4.2. By fiat, this lattice sits at $z_0 = k^m a$. The number of bonds between the two sites at depth m is $x_0/(k^m a)$ (see Fig. 4.2 for the case $k = 2$). It follows that the proper length of the line connecting the two points is just

$$\begin{aligned} |\gamma_1|_{\text{MERA}} &= L_1 \cdot (\text{number of bonds between endpoints}) \\ &= L_1 \left.\frac{x_0}{z_0}\right|_{z_0 = k^m a} . \end{aligned} \tag{4.5}$$

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

Similarly, consider two lattice sites on the UV-most lattice separated by a coordinate distance x_0. If we assume that $x_0 \gg a$, then the shortest path (geodesic) in the MERA connecting the two lattice sites is the path that goes up in the renormalization

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

$$\begin{aligned} |\gamma_2|_{\text{MERA}} &= L_2 \cdot (\text{number of bonds in the geodesic}) \\ &= 2L_2 \log_k \left(\frac{x_0}{a}\right) \end{aligned} \tag{4.6}$$

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

4.3.2 Limits on sub-AdS scale physics

One aspect of the matching of geodesics that is immediately apparent is that the MERA scales L_1 and L_2 that parametrize the proper distance between lattice sites are of order the AdS scale L or larger, as was also noted in Refs. [76, 85]. This runs counter to the typical expectation that, in a discretization of spacetime, one expects the granularity to be apparent on the UV, rather than the IR, scale. That is, sub-AdS scale locality is not manifested in the MERA construction and must be encoded within each tensor factor [86].

One could try to evade this difficulty by attempting to redefine the MERA coordinates $(x, z)^{\text{MERA}}$ (those of Fig. 4.2) as functions of the AdS coordinates $(x, z)^{\text{AdS}}$ (those of Fig. 4.3) and taking a continuum limit; above, we assumed that the two sets of coordinates were simply identified. That is, suppose $x^{\text{MERA}} = f(x^{\text{AdS}})$ and $z^{\text{MERA}} = g(z^{\text{AdS}})$. (For example, one could consider $f(x) = \varepsilon x$ for small ε and imagine taking the continuum limit, with the aim of making L_1 much smaller than the AdS scale.) If a is still the UV cutoff on the AdS side, then in the MERA we have $f(a)$ as the UV-most lattice spacing and $g(a)$ as the UV cutoff in the holographic direction. Consider the computation of $|\gamma_1|$. From the AdS side, we have $|\gamma_1|_{\text{AdS}} = Lx_0^{\text{AdS}}/z_0^{\text{AdS}}$. On the MERA side, the number of sites spanned by $x_0^{\text{MERA}} = f(x_0^{\text{AdS}})$ is $x_0^{\text{MERA}}/k^m f(a)$, while the holographic coordinate

is $z_0^{\mathrm{MERA}} = k^m g(a)$. Thus,

$$|\gamma_1|_{\mathrm{MERA}} = L_1 \frac{f(x_0^{\mathrm{AdS}})}{f(a)} \frac{g(a)}{g(z_0^{\mathrm{AdS}})}. \tag{4.7}$$

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

$$g(z_0^{\mathrm{AdS}}) \frac{\partial}{\partial x_0^{\mathrm{AdS}}} |\gamma_1| = L_1 \frac{f'(x_0^{\mathrm{AdS}})}{f(a)} g(a) = L \frac{g(z_0^{\mathrm{AdS}})}{z_0^{\mathrm{AdS}}}. \tag{4.8}$$

Since the right side of the first equality only depends on x_0^{AdS} and the second equality only depends on z_0^{AdS}, but we can vary both parameters independently, both expressions must be independent of both AdS coordinates. Therefore, we must have $f(x) = \varepsilon_x x$ and $g(z) = \varepsilon_z z$ for some constants ε_x and ε_z. Plugging everything back into Eq. (4.7) and comparing with $|\gamma_1|_{\mathrm{AdS}}$, we again find that $L_1 = L$, so no continuum limit is possible. Similarly, in computing $|\gamma_2|$, we note that the number of bonds between the endpoints on the UV-most lattice level is $x_0^{\mathrm{MERA}}/f(a)$, so the geodesic connecting the endpoints has $2\log_k(x_0^{\mathrm{MERA}}/\varepsilon_x a)$ bonds. On the other hand, we have $|\gamma_2|_{\mathrm{AdS}} = 2L\ln(x_0^{\mathrm{AdS}}/a) = 2L\ln(x_0^{\mathrm{MERA}}/\varepsilon_x a)$. That is, in equating $|\gamma_2|_{\mathrm{AdS}}$ and $|\gamma_2|_{\mathrm{MERA}}$, we must again set $L_2 = L\ln k$. We thus also find that no continuum limit is possible in the holographic direction. That is, we have shown that there is a constant normalization freedom in the definition of each of the coordinate distances on the AdS and MERA sides of any AdS/MERA duality, but such a coordinate ambiguity is unphysical and does not allow one to take a continuum limit. One still finds that the physical MERA parameters L_1 and L_2 are AdS scale. This means that there truly is no sense in which a discrete MERA can directly describe sub-AdS scale physics without the addition of supplemental structure to replace the individual tensors. This fact limits the ability of the MERA to be a complete description of the gravity theory without such additional structure. It might be the case that one needs a field theoretic generalization of the MERA, such as continuous MERA (cMERA) [88–90] or some local expansion of the individual tensors into discrete tensor networks with a different graph structure to describe sub-AdS physics, but such a significant generalization of the tensor network is beyond the scope of this work and in any case would no longer correspond to a MERA proper.

4.4 Constraints from Boundary Entanglement Entropy

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

4.4.1 MERA and CFT Entanglement Entropy

For a (1+1)-dimensional CFT in a pure state, the von Neumann entropy of a finite interval B, which is typically referred to as the entanglement entropy, is known to be [62, 63]

$$S(B) = \frac{c}{3} \ln \ell_0 \,, \tag{4.9}$$

where the length of the interval is much smaller than the system size. Here, ℓ_0 is the length of the interval in units of the UV cutoff. In the notation of the last section, we have $\ell_0 = x_0/a$. In the special case that the CFT is dual to AdS in $2+1$ dimensions, the central charge is set by the Brown–Henneaux formula [91],

$$c = \frac{3L}{2G}. \tag{4.10}$$

Also note that the length of the geodesic that connects the two ends of B (the curve γ_2 in Fig. 4.3) is given in Eq. (4.4) by $|\gamma_2| = 2L \ln \ell_0$. The Brown–Henneaux relation allows us to reproduce the Ryu–Takayanagi formula [17, 92] from the entanglement entropy,

$$S(B) = \frac{\text{area}(\tilde{B})}{4G} \,, \tag{4.11}$$

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

The MERA calculation of the entanglement entropy of ℓ_0 sites in the CFT has an analogous geometric interpretation. Suppose one is given the MERA representation of a lattice CFT ground state, i.e., one uses a MERA to generate the CFT state. Denote by $S_{\mathrm{MERA}}(\ell_0)$ the entanglement entropy of the resulting state restricted to ℓ_0 sites. In Ref. [87], it is shown that for a specific, optimal choice of ℓ_0 sites, for ℓ_0 parametrically large, the following bound is placed on $S_{\mathrm{MERA}}(\ell_0)$ for a MERA with $k = 2$:

$$S_{\mathrm{MERA}}(\ell_0) \leq 2 \log_2 \ell_0 \ln \chi. \tag{4.12}$$

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

The causal cone of a region B consisting of ℓ_0 contiguous UV sites in a MERA resembles a bulk extremal surface for the boundary region B. Given ℓ_0 sites in the UV, their causal cone is defined as the part of the MERA on which the reduced density matrix (or in other words, the state) of B depends. An example of a causal cone is illustrated in Fig. 4.4.

In particular, note that the number of bonds that a causal cone crosses up to any fixed layer scales like the length of the boundary of the causal cone up to that layer. It is in this sense that Eq. (4.12) is a MERA version of Ryu–Takayanagi. Also note that the width of the causal cone shrinks by a factor of $\sim 1/k$ after every renormalization step until its width is comparable to k. As such, if one denotes the width of the causal cone at a layer m by ℓ_m, then ℓ_m is roughly constant for all m greater than some $\bar{m}$ (see Fig. 4.4). The scale $\bar{m}$ is called the *crossover scale*.

For general k, it is also possible to formulate a bound similar to Eq. (4.12) for the entanglement entropy of ℓ_0 sites. For parametrically large ℓ_0, we find that

$$S_{\mathrm{MERA}}(\ell_0; B) \leq 4(k-1) \log_k \ell_0 \ln \chi \,. \tag{4.13}$$

We demonstrate this bound in Sec. 4.7 using techniques that are similar to those

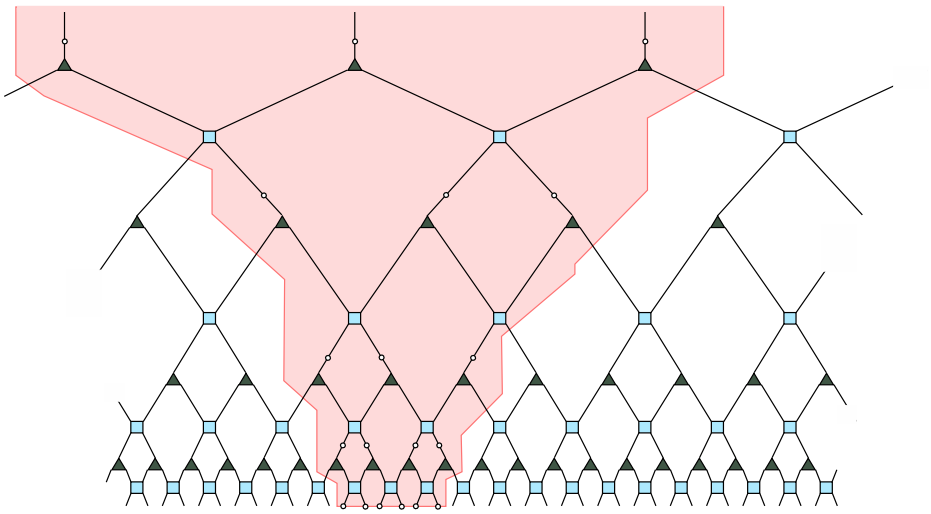

Figure 4.4: Causal cone (shaded) for a set of $\ell_0 = 6$ sites in a MERA with $k = 2$. The width ℓ_m of the causal cone at depth m is $\ell_1 = 4$, $\ell_2 = 3$, $\ell_3 = 3$, $\ell_4 = 3$, etc. The crossover scale for this causal cone occurs at $\bar{m} = 2$. Between the zeroth and first layer, $n_1^{\text{tr}} = 2$ bonds are cut by the causal cone. Similarly, $n_2^{\text{tr}} = 2$, $n_3^{\text{tr}} = 3$, etc.

developed in Ref. [87]. In particular, note that we do not allow ourselves to choose the location of the ℓ_0 sites in question. As such, we remind ourselves that S_{MERA} can depend on the location of the region B (and not only its size) by including it in the argument of S_{MERA}. This is also the reason why our Eq. (4.13) is more conservative than the optimal bound given in Eq. (4.12).

4.4.2 Constraining S_{MERA}

Let us examine Eq. (4.13) a bit more closely. As discussed in Sec. 4.7, $4(k-1)$ is an upper bound on the number of bonds that the causal cone could cut at any given depth m below the crossover scale $\bar{m}$. (The crossover scale $\bar{m}$ is attained after roughly $\log_k \ell_0$ renormalization steps.) For a given causal cone, i.e., for ℓ_0 sites at a given location with respect to the MERA, let us parametrize our ignorance by writing

$$S_{\text{MERA}}(\ell_0; B) \leq 4 f_B(k) \log_k \ell_0 \ln \chi \,, \tag{4.14}$$

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

$$f_B(k) \equiv \frac{1}{4\bar{m}} \sum_{m=0}^{\bar{m}-1} n_m^{\mathrm{tr}}\,, \tag{4.15}$$

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

Each cut bond contributes at most $\ln \chi$ to the entropy (the case of maximal entanglement). As such, it is instructive to introduce a parameter $\eta_B \in [0,1]$ that describes the degree of entanglement of the sites in the causal cone. In doing so we may rewrite the inequality (4.14) as an equality:

$$S_{\mathrm{MERA}}(\ell_0; B) = 4 f_B(k) \log_k \ell_0 \cdot \eta_B \ln \chi. \tag{4.16}$$

The quantity $\eta_B \ln \chi$ is the average entanglement entropy per cut bond in the causal cone of B. Equivalently, Eq. (4.16) may be taken as the definition of η_B.

This definition of η_B of course depends on the location of B and only applies to bonds that are cut by the causal cone of B. In what follows, it will be advantageous to have a notion of average entanglement entropy per bond that applies to *all* bonds in the MERA. To this end, start with a lattice consisting of ℓ_{tot} sites in total and consider the limit in which the size of a region B is unbounded but where the ratio $\ell_0/\ell_{\mathrm{tot}}$ is held constant (so that B does not grow to encompass the whole domain of the CFT). In this limit, $S_{\mathrm{MERA}}(\ell_0; B) \to S_{\mathrm{MERA}}(\ell_0)$ and $f_B(k) \to f(k)$ should be independent of the exact location of B, i.e., S_{MERA} should exactly agree with Eq. (4.9). Let us consequently define the average entanglement entropy per bond in the MERA:

$$\eta \ln \chi = \lim_{\ell_0 \to \infty} \frac{S_{\mathrm{MERA}}(\ell_0)}{4 f(k) \log_k(\ell_0)}\,. \tag{4.17}$$

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

Intuitively, one would not expect each individual bond in the MERA to be maximally entangled and so it should be possible to constrain η more tightly than $\eta \leq 1$. This expectation is made more precise via the following considerations. To begin, consider a MERA with $k = 2$ and examine a pair of isometries at a fixed depth

m. As indicated in Fig. 4.5 (a), let ρ_2 denote the density matrix of the bonds and ancillae emanating from the two isometries and let ρ_1 denote the density matrix of the four highlighted bonds below the isometries. We again assume that the ancillae are initialized to the pure product state composed of factors of $|0\rangle$. Taking into account the ancillae, or in other words promoting the isometries to unitaries, we see that ρ_1 and ρ_2 are related by a unitary transformation, so $S(\rho_1) = S(\rho_2)$. By assumption, the state of each ancilla is $|0\rangle$, so $\rho_2 = \tilde{\rho}_2 \otimes |0\rangle\langle 0| \otimes |0\rangle\langle 0|$ for some density matrix $\tilde{\rho}_2$. This in turn implies that $S(\rho_2) = S(\tilde{\rho}_2) \leq 2 \ln \chi$. From the definition of η above, the entanglement entropy of a single bond is asymptotically given by $\eta \ln \chi$, so $S(\rho_1) \simeq 4\eta \ln \chi$. It therefore follows that $\eta \leq 1/2$.

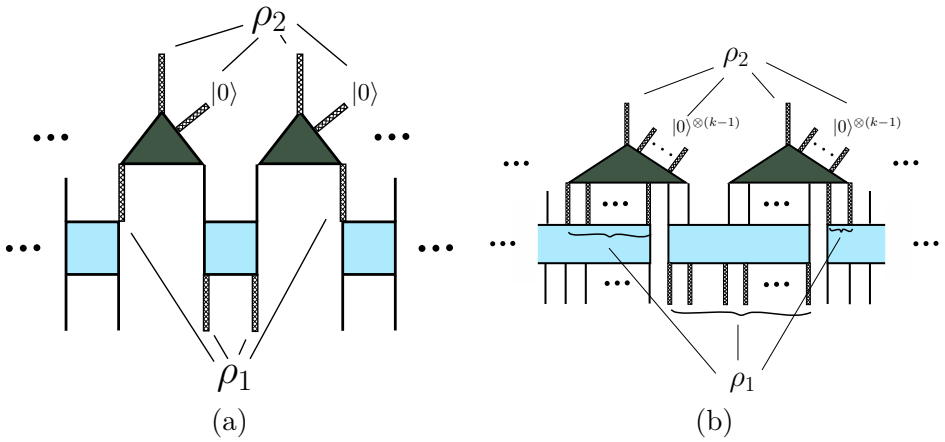

Figure 4.5: A pair of isometries with their ancillae explicitly indicated for a MERA with (a) $k = 2$ and (b) general k. The thick bonds below the isometries, the state of which is denoted by ρ_1, are unitarily related to the bonds that exit the isometries and the ancillae, the state of which is denoted by ρ_2.

For general k, the argument is nearly identical. We again begin by considering a pair of isometries at a given level m (see Fig. 4.5 (b)). Analogously with the $k = 2$ case, let ρ_2 denote the density matrix of the two bonds and $2k - 2$ ancillae emanating from the two isometries and let ρ_1 denote the density matrix of the $2k$ highlighted bonds below the isometries. There is only one disentangler that straddles both of the isometries in question for any layout of the MERA. As such, at most k of the lower bonds enter a disentangler from below and the rest directly enter the isometries. Here as well ρ_1 and ρ_2 are related by a unitary transformation so that

$S(\rho_1) = S(\rho_2)$. Similarly, $\rho_2 = \tilde{\rho}_2 \otimes (|0\rangle\langle 0|)^{\otimes 2k-2}$ for some density matrix $\tilde{\rho}_2$, so $S(\rho_2) = S(\tilde{\rho}_2) \leq 2 \ln \chi$. The region described by ρ_1 always consists of $2k$ bonds, so we may again asymptotically write $S(\rho_1) \simeq 2k\eta \ln \chi$. It therefore follows that $k\eta \leq 1$, and since $f(k) \leq (k-1)$, we may write

$$\eta f(k) \leq \frac{k-1}{k} \,. \tag{4.18}$$

We note that, in computational practice, one typically does not use the "worst-case scenario" construction explored in Sec. 4.7; a more conventional construction would result in a tighter bound on $f(k)$ and hence a stricter inequality than Eq. (4.18). For our purposes, however, we will remain as conservative as possible and therefore use the inequality (4.18) in our subsequent bounds.

4.4.3 Matching to the CFT

Finally, we obtain a constraint on k, χ, and η in terms of the central charge c by collecting the results of this section. Let us work in the limit where the interval is much larger than the lattice spacing, $\log_k \ell_0 \gg 1$. We have seen that this is precisely the regime in which η and $f(k)$ are well-defined quantities independent of the choice of B. It is also the regime in which we can equate the CFT entropy $S(\ell_0) = (c/3) \ln \ell_0$ with the MERA entropy (4.16). Doing so, the central charge is given by

$$c = \frac{3L}{2G} = 12\eta \, f(k) \frac{\ln \chi}{\ln k} \,. \tag{4.19}$$

Then in light of Eq. (4.18), we find that

$$c \leq 12 \left(\frac{k-1}{k \ln k} \right) \ln \chi \,. \tag{4.20}$$

To recapitulate, given a CFT with central charge c and a MERA representation of its ground state, a necessary condition for a consistent AdS/MERA correspondence is that the MERA parameters k and χ satisfy the constraint (4.20). Importantly, this implies that, for a well-defined semiclassical spacetime (for which $c \gg 1$), the bond dimension χ must be exponentially large in the size of the AdS scale compared to the Planck scale.

Let us also note that we can still obtain a bound from Eq. (4.19), albeit a weaker one, without using the result of Eq. (4.18). Recall that this latter result relies on having unentangled ancillae in the MERA. This is not necessarily the case for other tensor network bulk constructions, as we will subsequently discuss. As such, if we disregard the result of Eq. (4.18), we still have by virtue of their definitions that $f(k) \leq k - 1$ and $\eta \leq 1$. The following weaker but more general bound on the central charge therefore follows from Eq. (4.19) for such generalized tensor networks:

$$c \leq 12 \left(\frac{k-1}{\ln k} \right) \ln \chi. \tag{4.21}$$

4.5 Constraints from Bulk Entanglement Entropy

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

4.5.1 The Bousso Bound

The notion of placing bounds on the entropy of regions of spacetime in a quantum gravity theory has been explored for many years, first in the context of black hole thermodynamics [93] and the Bekenstein bound [94] and later in more general holographic contexts, culminating in the covariant entropy bound, i.e., the Bousso bound [95, 96].

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

Let our spacelike surface $\mathcal{B}$ be a 2-ball of area A on a spacelike slice of AdS and choose as the lightsheet the ingoing future-directed null congruence. This lightsheet will sweep out the entire interior of $\mathcal{B}$ and will terminate at a caustic at the center of $\mathcal{B}$. Since the system is static, the entropy S passing through this lightsheet is the entropy of the system on $\mathcal{B}$, which by the Bousso bound satisfies

$$S(\mathcal{B}) \leq \frac{A}{4G}\,. \tag{4.22}$$

It is natural to cast the Bousso bound as a constraint on the dimension of the bulk Hilbert space. As argued in Ref. [97], the thermodynamic entropy of a system about which we only know the boundary area A is just the logarithm of the dimension of the true Hilbert space of the bulk region in question (as opposed to the naïve Hilbert space in quantum field theory), which the Bousso bound implies is less than $A/4G$.[1] As such, if we denote the Hilbert space of $\mathcal{B}$ by $\mathcal{H}_{\mathcal{B}}$, let us replace Eq. (4.22) with the slightly more concrete statement

$$\ln \dim \mathcal{H}_{\mathcal{B}} \leq \frac{A}{4G}. \tag{4.23}$$

4.5.2 A MERA version of the Bousso Bound

Our aim is to compute both sides of the inequality (4.23) using the MERA. For this calculation, it is instructive to change our parametrization of the hyperbolic plane from coordinates (x, z), which take values in the half-plane $z > 0$, to coordinates (ρ, θ), which take values in a disk $0 \leq \rho < 1$, $0 \leq \theta < 2\pi$. Embeddings of the MERA in a disk are often depicted in the literature, e.g., [100]; here we make this coordinate transformation explicit, since we wish to carefully study the geometric properties of the MERA.

To begin, consider a MERA consisting of a single tree that contains a finite number of layers m. This situation is illustrated in Fig. 4.6 (a) for $k = 2$ and $m = 4$. Note

[1] Moreover, it is known that there exists an asymptotically-AdS bulk configuration that saturates the Bousso bound, namely, the BTZ black hole [70, 98], which further implies that $\ln \dim \mathcal{H}_{\mathcal{B}}$ in fact equals $A/4G$. However, we will not need this stronger assertion in what follows. A similar but unrelated result equating the area of a region with its entanglement entropy in vacuum was obtained in Ref. [99].

that such a MERA begins with a top-level tensor at the m^{th} level that seeds the rest of the MERA in the IR.

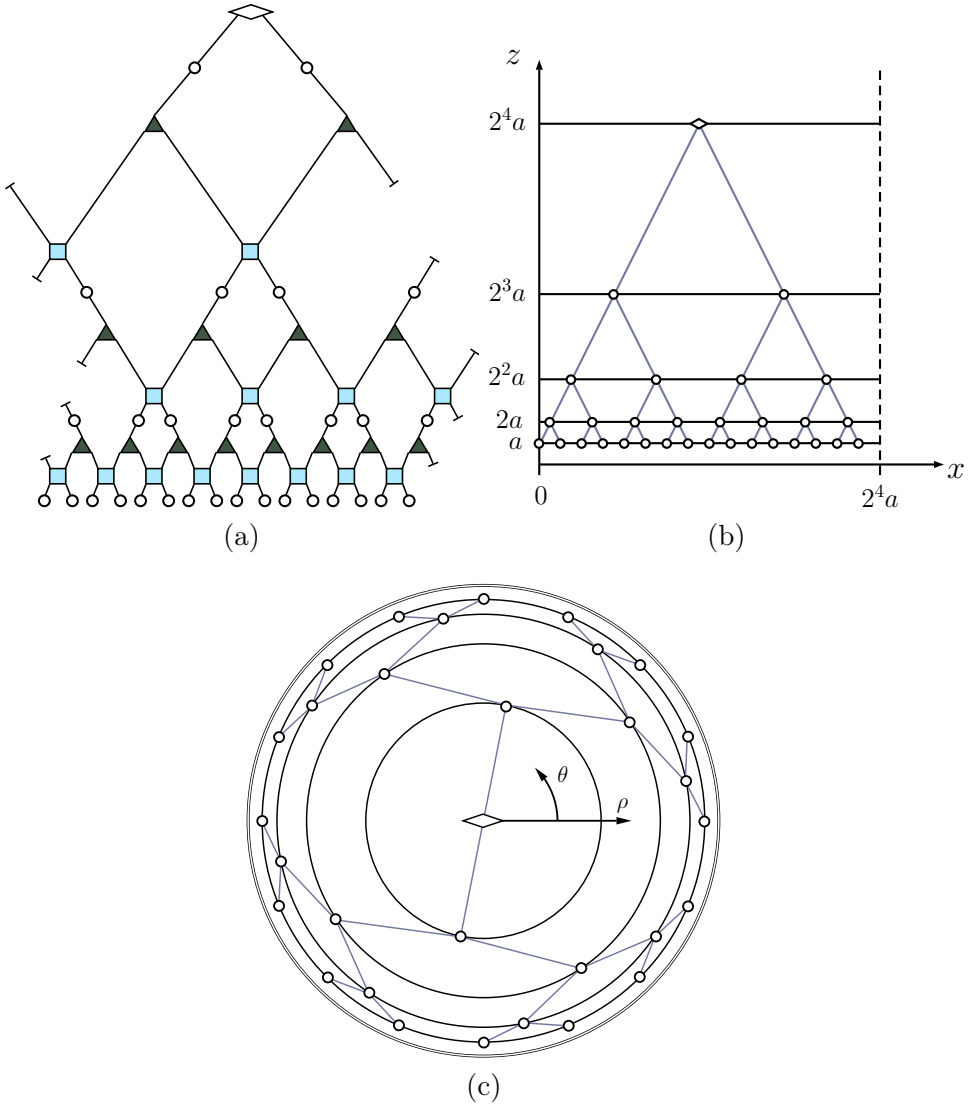

Figure 4.6: (a) A $k = 2$ MERA consisting of $m = 4$ layers and with periodic boundary conditions, (b) the corresponding embedding in (x, z) coordinates, and (c) the embedding in (ρ, θ) coordinates.

The base of the MERA is made up of k^m sites. Without loss of generality, let us locate the leftmost site of the base of the MERA at $x = 0$, so that the UV-most sites sit at coordinates $(x, z) = (na, a)$, where $n = 0, 1, 2, \ldots, (k^m - 1)$ as shown in Fig. 4.6 (b). Let us also assume periodic boundary conditions for this MERA and hence identify $x = 0$ and $x = k^m a$.

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

$$\begin{aligned} \rho &= \frac{k^m a - z}{k^m a}, \\ \theta &= 2\pi \frac{x}{k^m a}. \end{aligned} \tag{4.24}$$

In these coordinates, the metric reads

$$\mathrm{d}s^2 = \frac{L^2}{(1-\rho)^2} \left[\mathrm{d}\rho^2 + \left(\frac{\mathrm{d}\theta}{2\pi} \right)^2 \right], \tag{4.25}$$

cf. Eq. (4.3). This embedding of the MERA is shown in Fig. 4.6 (c); the top-level tensor always sits at $\rho = 0$ and the lower layers of the MERA are equally spaced on circles of radii 1/2, 3/4, 7/8, ... that are centered at $\rho = 0$.

More generally, one could construct a top-level tensor that has T legs, each of which begets a tree of sites. In this case, $x = 0$ and $x = Tk^{m-1}a$ are identified, so one should define the angular variable as $\theta \equiv 2\pi x/(Tk^{m-1}a)$. The metric (4.25) is correspondingly modified and reads

$$\mathrm{d}s^2 = \frac{L^2}{(1-\rho)^2} \left[\mathrm{d}\rho^2 + \frac{T^2}{k^2} \left(\frac{\mathrm{d}\theta}{2\pi} \right)^2 \right]. \tag{4.26}$$

This situation is depicted in Fig. 4.7. (If $T = k$, however, then it is not necessary to introduce any new structure in addition to the disentanglers and isometries that were already discussed, i.e., one may take the top-level tensor to be one of the isometries.)

We may immediately compute the right-hand side of Eq. (4.23). Let the ball $\mathcal{B}$ be centered about $\rho = 0$, and suppose $\mathcal{B}$ contains the top-level tensor, the sites at the top tensor's legs, and then the first $N_{\mathcal{B}}$ generations of the MERA emanating from these sites, as indicated in Fig. 4.7. The boundary of $\mathcal{B}$ is a circle at constant ρ, so its circumference according to the MERA is $A = Tk^{N_{\mathcal{B}}}L$. As such, we may write

$$\frac{A}{4G} = \frac{Tk^{N_{\mathcal{B}}}L}{4G} = \frac{Tk^{N_{\mathcal{B}}}c}{6}, \tag{4.27}$$

where in the second equality we used the Brown-Henneaux relation, Eq. (4.10).

How one evaluates the left-hand side of Eq. (4.23) using the MERA is not as immediate. Recall that $\mathcal{H}_{\mathcal{B}}$ is the Hilbert space of *bulk states*. The MERA, however, does

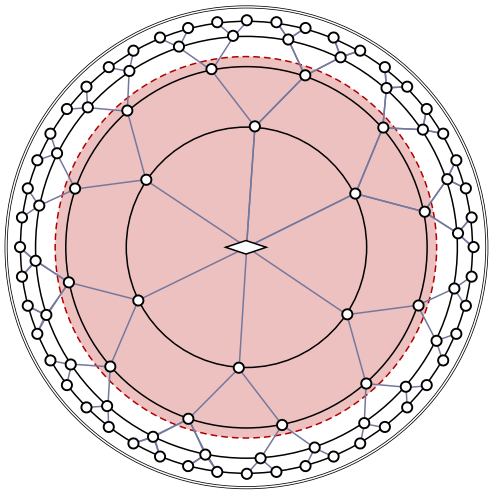

Figure 4.7: Disk parametrization of the Poincaré patch of AdS in which a MERA has been embedded. The top tensor of the MERA shown has $T = 6$. The shaded region is a ball $\mathcal{B}$, which is this case contains $N_{\mathcal{B}} = 1$ generation.

not directly prescribe the quantum-gravitational state in the bulk; it is not by itself a bulk-boundary dictionary. As we mentioned in Sec. 4.2.2, the minimal assumption that one can make is to posit the existence of a bulk Hilbert space factor V_{bulk} associated with each MERA site that is not located at the top tensor. To keep the assignment general, we assign a factor V_{T} to the top tensor. The dimensionality of each V_{bulk} factor should be the same in order to be consistent with the symmetries of the hyperbolic plane. The assumption of a Hilbert space factor at every MERA site is minimal in the sense that it introduces no new structure into the MERA. A true AdS/MERA correspondence should dictate how states in the bulk Hilbert space are related to boundary states. However, for our analysis, it is enough to simply postulate the existence of the bulk Hilbert space factors V_{bulk} and V_{T}, each of which may be thought of as localized to an AdS-scale patch corresponding to the associated MERA site.

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

$$\mathcal{N}_{\mathcal{B}} = T \sum_{i=0}^{N_{\mathcal{B}}} k^i = T \left(\frac{k^{N_{\mathcal{B}}+1} - 1}{k - 1} \right) . \tag{4.28}$$

As such, the Hilbert space of bulk states restricted to $\mathcal{B}$ is $\mathcal{H}_{\mathcal{B}} = (V_{\text{bulk}})^{\otimes \mathcal{N}_{\mathcal{B}}} \otimes V_{\text{T}}$.

Next, suppose that $\dim V_{\text{bulk}} = \tilde{\chi}$ and that $\dim V_{\text{T}} = \tilde{\chi}_{\text{T}}$, where, like χ, $\tilde{\chi}$ and $\tilde{\chi}_{\text{T}}$ are some fixed, $N_{\mathcal{B}}$-independent numbers. Then $\dim \mathcal{H}_{\mathcal{B}} = \tilde{\chi}_{\text{T}}(\tilde{\chi}^{\mathcal{N}_{\mathcal{B}}})$. Note that one would expect χ and $\tilde{\chi}$ to have a very specific relationship in a true bulk/boundary correspondence, the nature of which will be explored later in this section. Combining Eqs. (4.27) and (4.28), the dimensionality of $\mathcal{H}_{\mathcal{B}}$ is upper bounded as follows:

$$\ln \dim \mathcal{H}_{\mathcal{B}} \leq \frac{A}{4G} \quad \Longrightarrow \quad T\left(\frac{k^{N_{\mathcal{B}}+1}-1}{k-1}\right)\ln\tilde{\chi} + \ln\tilde{\chi}_{\text{T}} \leq \frac{Tk_{\mathcal{B}}^{N}c}{6}\,. \tag{4.29}$$

After isolating c in Eq. (4.29) and using the result of Eq. (4.19), we find that

$$c = 12\eta f(k)\frac{\ln\chi}{\ln k} \geq 6\left(\frac{k^{N_{\mathcal{B}}+1}-1}{k^{N_{\mathcal{B}}}(k-1)}\ln\tilde{\chi} + \frac{1}{Tk^{N_{\mathcal{B}}}}\ln\tilde{\chi}_{\text{T}}\right). \tag{4.30}$$

Next, let us consider this inequality in the limit of large $N_{\mathcal{B}}$. A motivation for this limit is the fact that the natural scale of validity of an AdS/MERA correspondence is super-AdS, as was established in Sec. 4.3. Moreover, by virtue of its definition, there is always an ambiguity of order the AdS scale in the radius of the ball $\mathcal{B}$. That is, the region in AdS denoted by $\mathcal{B}$ is only well-defined in the MERA if $\mathcal{B}$ is large compared to the AdS scale L. Taking the limit of large $N_{\mathcal{B}}$, Eq. (4.30) reduces to

$$\eta f(k) \geq \frac{k\ln k}{2(k-1)}\left(\frac{\ln\tilde{\chi}}{\ln\chi}\right). \tag{4.31}$$

By using the bound on $\eta f(k)$ given by Eq. (4.18), we arrive at a constraint on k, χ, and $\tilde{\chi}$:

$$\frac{k^2\ln k}{2(k-1)^2}\left(\frac{\ln\tilde{\chi}}{\ln\chi}\right) \leq 1. \tag{4.32}$$

In principle, the above inequality could be satisfied for any k, provided that the dimension $\tilde{\chi}$ of the factors V_{bulk} can be arbitrarily chosen with respect to the bond dimension χ. However, the essence of holography, that the bulk and boundary are dual descriptions of the same degrees of freedom and therefore have isomorphic Hilbert spaces [73], implies a relation between χ and $\tilde{\chi}$. Namely, for a MERA with a total of N levels of sites in the bulk strictly between the UV-most level and the top-level tensor, the number of bulk sites $\mathcal{N}_{\text{bulk}}$ that are not located at the top tensor is given by Eq. (4.28) with $N_{\mathcal{B}} = N$, and the number of sites in the boundary

description is $\mathcal{N}_{\text{boundary}} \equiv Tk^{N+1}$. The bulk Hilbert space thus has dimension $\tilde{\chi}^{\mathcal{N}_{\text{bulk}}}\tilde{\chi}_{\text{T}}$ and the boundary Hilbert space has dimension $\chi^{\mathcal{N}_{\text{boundary}}}$. Equating[2] the dimension of the bulk and boundary Hilbert spaces then yields

$$\frac{\ln\tilde{\chi}}{\ln\chi} = \frac{1}{\mathcal{N}_{\text{bulk}}}\left(Tk^{N+1} - \frac{\ln\tilde{\chi}_{\text{T}}}{\ln\chi}\right) \overset{N\,\text{large}}{\to} k-1, \tag{4.33}$$

where we took the limit of N large, consistent with Eq. (4.31) and in keeping with the expectation that the UV cutoff be parametrically close to the boundary at $\rho = 1$. Putting together Eqs. (4.32) and (4.33), we obtain a constraint on k alone:

$$\frac{k^2\ln k}{2(k-1)} \leq 1. \tag{4.34}$$

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

If we relax this bound and, instead of Eq. (4.18), only observe the weaker, natural bounds $\eta \leq 1$ and $f(k) \leq k-1$ as discussed at the end of Sec. 4.4.3, the constraint (4.34) is correspondingly modified:

$$\frac{k\ln k}{2(k-1)} \leq 1. \tag{4.35}$$

In contrast to Eq. (4.34), this latter bound can be satisfied, but only for $k = 2$, 3, or 4. As such, other AdS/tensor network correspondences, in which the ancillae are perhaps entangled and therefore do not describe a conventional MERA, are not ruled out. Note that we never needed to compute bulk entanglement entropy explicitly —

[2] We recognize that there are other proposals [26, 31] that do not require an exact equivalence between the bulk and boundary Hilbert spaces, but, even in these cases, there is the requirement of an exact equivalence between the logical qubits on the boundary with the Hilbert space of the bulk.

and therefore did not need to treat separately the possibility of entanglement among ancillae — because we cast the Bousso bound as a constraint on the size of the bulk Hilbert space itself. The appearance of η in Eq. (4.31) corresponds to entanglement in the boundary theory as computed by the tensor network; Eqs. (4.31) and (4.33) still apply.

4.6 Conclusion

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

1. In matching the discrete graph geometry of the MERA to the continuous geometry of a spatial slice of AdS, we demonstrated that the MERA describes geometry only on scales larger than the AdS radius. Concretely, as shown in Sec. 4.3, the proper length assigned to the spacing between adjacent sites in the MERA lattice must be the AdS scale.

2. By requiring that the entropy of a set of boundary sites in the MERA — whose computation is a discrete realization of the Ryu–Takayanagi formula — be equal to the CFT ground state entropy of the same boundary region in the thermodynamic limit, we obtained a constraint on the parameters that describe a MERA in terms of the CFT central charge [Eqs. (4.20) and (4.21)], which implies that the bond dimension χ must be exponentially large in the ratio of the AdS scale to the Planck scale.

3. In the natural construction of a bulk Hilbert space ($\mathcal{H}_{\text{bulk}}$) using the MERA, we used the Bousso bound to constrain the dimension of $\mathcal{H}_{\text{bulk}}$. When combined with our previous results, we found that any strict AdS/MERA correspondence cannot satisfy the resulting constraint, Eq. (4.34). Upon relaxing the definition

of the MERA or allowing for additional structure, however, we obtained a looser constraint, Eq. (4.35), which may not rule out some other AdS/tensor network correspondences.

In particular, more general correspondences between AdS and MERA-like tensor networks, in which we allow the ancillae to be entangled when reproducing the CFT ground state [and for which Eq. (4.35) applies in place of Eq. (4.34)] are not ruled out by our bounds, provided that the rescaling factor $k = 2$, 3, or 4. Further, it is interesting to note that our bounds extend to states other than the vacuum that are described by a MERA. One such example, namely, states at finite temperature dual to black holes in AdS, is discussed in Sec. 4.8 below.

While the consistency conditions that we found are specific to the MERA tensor network, many of the ideas and techniques that we used apply equally well to other tensor networks. In the EHM, for instance, the type of bulk Hilbert space dimensionality arguments that we made based on the covariant entropy bound may be directly transferred to the EHM. The same stringent final constraints that we derived do not apply to the EHM, however, since it is unclear to what extent the EHM reproduces the Ryu–Takayanagi formula (which renders the results of Sec. 4.4 inapplicable). Our bulk Hilbert space arguments similarly apply to the holographic error-correcting code proposal in Ref. [31], which furthermore purports to reproduce a version of the Ryu–Takayanagi formula. It is presently unknown, however, whether the boundary state of a holographic code can represent the ground state of a CFT, so an identification of entropies similar to the identification $S_{\mathrm{MERA}} = S_{\mathrm{CFT}}$, upon which our boundary entropy constraints so crucially depend, cannot yet be made.

In closing, we have found several consistency conditions that any AdS/MERA correspondence must satisfy. The totality of these constraints rules out the most straightforward construal of an AdS/MERA correspondence. Other interesting holographic correspondences that are described by tensor networks more general than the MERA and that respect all of our bounds may indeed be possible. Our consistency conditions are nice validity checks for these correspondences when applicable and in

other cases they may inspire similar consistency conditions. The program of identifying the emergence of spacetime from the building blocks of quantum information is an ambitious one; stringent consistency conditions, such as those presented in this paper, are important for elucidating the subtleties in this quest and in providing guidance along the way.

Acknowledgements

We thank Bartek Czech, Glen Evenbly, Daniel Harlow, Shamit Kachru, Aleksander Kubica, Shaun Maguire, Spiros Michalakis, Don Page, John Preskill, Bogdan Stoica, James Sully, Brian Swingle, and Guifré Vidal for helpful discussions. This research was supported in part by DOE grant DE-SC0011632 and by the Gordon and Betty Moore Foundation through Grant 776 to the Caltech Moore Center for Theoretical Cosmology and Physics. N.B. was supported by the DuBridge postdoctoral fellowship at the Walter Burke Institute for Theoretical Physics. A.C.-D. and C.C. were supported by the NSERC Postgraduate Scholarship program. G.N.R. was supported by a Hertz Graduate Fellowship and a NSF Graduate Research Fellowship under Grant No. DGE-1144469.

4.7 Entropy bound for general MERAs

Following the method presented in Ref. [87], let us compute an upper bound for the entanglement entropy of a region B consisting of ℓ_0 sites in a MERA with rescaling factor k. We will use the notation of Ref. [87] throughout.

First, recall the result from Ref. [87] that the entanglement entropy of a region consisting of ℓ_0 sites is bounded by

$$S_{\mathrm{MERA}}(\ell_0; B) \leq (\ell_{m'} + N^{\mathrm{tr}}_{m'}) \ln \chi. \tag{4.36}$$

The quantity $\ell_{m'}$ is the width of the causal cone at depth m' and $N^{\mathrm{tr}}_{m'} = \sum_{m=0}^{m'-1} n^{\mathrm{tr}}_m$ is the total number of sites that are traced out along the boundary of the causal cone. In other words, $N^{\mathrm{tr}}_{m'}$ is the number of bonds that are cut by the causal cone up to a depth m' (*cf.* Fig. 4.4). The quantity $\ln \chi$ is the maximum entanglement

entropy that each site that is traced out could contribute to $S_{\mathrm{MERA}}(\ell_0; B)$. Note that Eq. (4.36) holds for all $m' \geq 0$.

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

Nevertheless, we can compute an upper bound for $S_{\mathrm{MERA}}(\ell_0; B)$ by considering a worst-case scenario for the number of bonds cut by the causal cone. We begin by asking: What is the largest number of bonds that a causal cone could cut in one renormalization step at a depth m'? The layout of disentanglers and isometries that produces this situation is shown at one side of a causal cone in Fig. 4.8. If the causal cone at the bottom of the renormalization step incorporates a single bond that goes into a disentangler accepting k bonds, then the causal cone must cut the other $k-1$ bonds entering the disentangler. Then if this disentangler is arranged so that its leftmost outgoing bond is the first bond to enter an isometry from the right, the causal cone must cut the other $k-1$ bonds entering the isometry. If this arrangement is mirrored on the other side of the causal cone, we see that $4(k-1)$ bonds are cut by the causal cone in this renormalization step, i.e., $n_{m'}^{\mathrm{tr}} = 4(k-1)$.

Recall that for any finite ℓ_0, after a fixed number of renormalization steps, the width of the causal cone remains constant for any further coarse-grainings. The depth at which this occurs is called the crossover scale and is denoted by $\bar{m}$. Therefore, the causal cone will cut the largest possible number of bonds when the arrangement described above and depicted in Fig. 4.8 occurs at every step up until the crossover

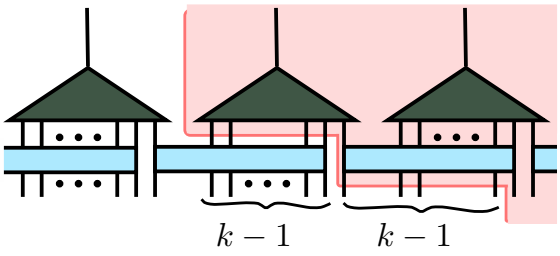

Figure 4.8: Left side of a causal cone that cuts the maximum possible number of bonds over the course of one renormalization step. The rectangles are disentanglers that accept k bonds as input and the triangles are isometries that coarse-grain k bonds into one. The causal cone is the shaded region. If this situation is mirrored on the right side of the causal cone, then $4(k-1)$ bonds are cut in this renormalization step.

scale. Then, by Eq. (4.36), the entropy bound is given by

$$S_{\mathrm{MERA}}(\ell_0; B) \leq (\ell_{\bar{m}} + 4(k-1)\bar{m}) \ln \chi, \tag{4.37}$$

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

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

$$\ell_{m'} \;\leq\; \frac{\ell_0}{k^{m'}} + 5 \sum_{m=1}^{m'} \frac{1}{k^m} \;\leq\; \frac{\ell_0}{k^{m'}} + 5\,. \tag{4.38}$$

Setting $m' = \bar{m}$, it trivially follows that the crossover scale obeys $\bar{m} \leq \log_k \ell_0$. Furthermore, we notice that this is the scale at which the entanglement entropy is minimized if we trace over the remaining sites. In other words, the number of bonds cut by going deeper into the renormalization direction is no less than the bonds cut horizontally, so $4(k-1) \geq \ell_{\bar{m}}$ [3]. Applying the bounds for $\bar{m}$ and $\ell_{\bar{m}}$ on Eq. (4.37),

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

we arrive at an upper bound on $S_{\mathrm{MERA}}(\ell_0; B)$ for a k-to-one MERA,

$$S_{\mathrm{MERA}}(\ell_0; B) \leq 4(k-1)(1+\log_k \ell_0) \ln \chi. \tag{4.39}$$

When ℓ_0 is parametrically large, we neglect the $\mathcal{O}(1)$ contribution to the bound on $S_{\mathrm{MERA}}(\ell_0; B)$, which yields Eq. (4.13).

4.8 BTZ Black Holes and Thermal States in AdS/MERA

Thus far, we have found constraints on the structure of a MERA that can describe CFT states dual to the AdS_3 vacuum. One might ask whether these results extend to other constructions that exist in three-dimensional gravity. Although pure gravity in AdS_3 has no local or propagating degrees of freedom, there exist interesting non-perturbative objects, namely, BTZ black holes [70]. In this section, we extend our constraints on boundary entanglement entropy to these objects.

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

$$\mathrm{d}s^2 = -\frac{(r^2 - r_+^2)}{L^2}\mathrm{d}t^2 + \frac{L^2}{(r^2 - r_+^2)}\mathrm{d}r^2 + r^2 \mathrm{d}\phi^2 \,, \tag{4.40}$$

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

$$S_{\mathrm{BH}} = \frac{\mathrm{Area}}{4G} = \frac{\pi r_+}{2G} \,. \tag{4.41}$$

Let us now consider applying a MERA with rescaling factor k and bond dimension χ to a CFT at a finite temperature, where instead of minimizing the energy of the boundary state, one minimizes the free energy. In the CFT, turning on a temperature introduces a scale, going as the inverse temperature, which screens long-range correlations. Thus, the state will have classical correlations in addition to entanglement and the effect of a finite temperature on the entanglement entropy is the appearance of an extensive contribution. As one runs the MERA and coarse-grains, the thermal correlations that cannot be removed become more relevant. The MERA,

which is unable to remove the extensive contribution, truncates at a level with multiple sites. The schematic entanglement renormalization process is illustrated in Fig. 4.9. The state at the top level effectively factorizes, where each factor appears maximally mixed [76, 86]. A tractable realization of this tensor network structure recently appeared in Ref. [101], which found a MERA representation of a thermal state.

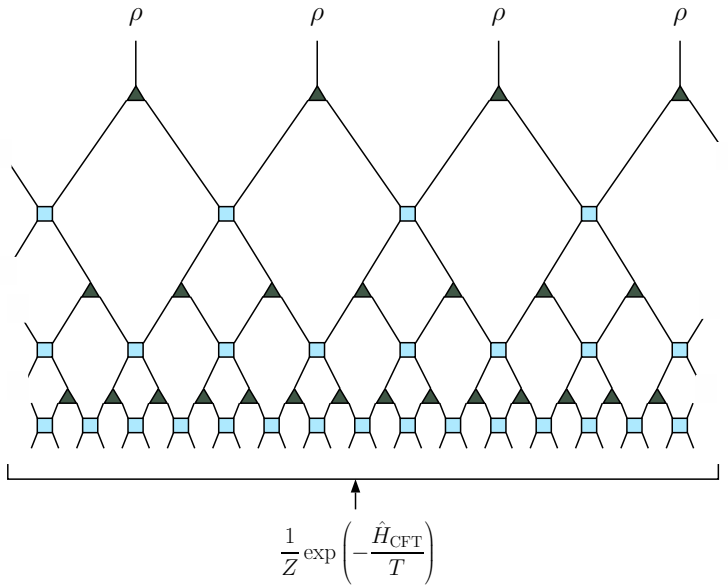

Figure 4.9: The MERA, when applied to a thermal CFT state $Z^{-1}\exp(-\hat{H}_{\mathrm{CFT}}/T)$, where $Z = \mathrm{tr}(\exp(-\hat{H}_{\mathrm{CFT}}/T))$, truncates after a finite number of layers. The boundary state at the top of the truncated MERA effectively factorizes into a product of maximally mixed states $\rho = I/\chi$.

Keeping in mind that the holographic dual of a finite-temperature state in the CFT is a black hole in AdS, where the temperature of the CFT corresponds to the Hawking temperature of the black hole, we note that the truncated MERA is suggestive of a black hole horizon [76]. If the MERA is to be interpreted as a discretization of the geometry, then the geometry has ended at some scale. Also, as we approach the horizon, the amount of Hawking radiation that we see increases and the temperature measured by an observer at the horizon diverges. The density matrix of some system in the infinite-temperature limit is given by the product of a maximally mixed state at each site, just like the state at the top of the MERA. It is important to note that,

as was pointed out in Ref. [101], in order to reproduce the correct thermal spectrum of eigenvalues, a small amount of entanglement must be present between the sites at the horizon. If the bond dimension were taken to be infinite, then the sites at the horizon truly would factorize. But for a finite bond dimension, one should really think of the horizon as a high-temperature state, with sites effectively factorized.

For small regions on the boundary, the length of the subtending bulk geodesic is subextensive and so the Ryu–Takayanagi formula maintains that the boundary region's entanglement entropy is subextensive as well. However, if we consider a large enough region on the boundary, the geodesic will begin to probe the horizon of the black hole. The geodesic will run along the black hole horizon and pick up an extensive contribution to the entropy. We consider a boundary theory living on a lattice consisting of n_{b} sites, with total system coordinate length $x_{\mathrm{sys}} = n_{\mathrm{b}} a$. In the limit as r approaches the boundary in the metric (4.40), we see that $T x_{\mathrm{sys}} = r_+/L$, as was pointed out in Refs. [17, 92]. We further note that this implies that the system coordinate size is of order AdS radius, $x_{\mathrm{sys}} = 2\pi L$.

Let us now view the MERA of Fig. 4.9 as a discretization of a BTZ spacetime and repeat the analysis of Sec. 4.3. In this discretization, the layers of the MERA lie along circles of fixed radius r in the coordinates of Eq. (4.40). Again, we ask what proper length L_1 separates sites in any given layer of the MERA.

First, note that a path at fixed r_0 that subtends an angle ϕ_0 has proper length $r_0\phi_0$. At the boundary of the MERA, we consider a region defined by $0 \leq \phi \leq \phi_0 = 2\pi x_0/x_{\mathrm{sys}}$, where x_0 is the coordinate length of the interval, consisting of ℓ_0 lattice sites. The boundary of the MERA is at a fixed radius $r = r_{\mathrm{b}}$. Naturally, the boundary radius r_{b} can be interpreted as a UV cutoff and is related to the lattice spacing a by $r_{\mathrm{b}} = L^2/a$ [17]. By equating the proper distance of the region in the MERA, $\ell_0 L_1$, with that at the boundary of the BTZ spacetime, $r_{\mathrm{b}}\phi_0$, we find the proper length between horizontal bonds to be $L_1 = L$.

With the foresight that the top of the MERA is suggestive of a black hole horizon with proper length $2\pi r_+$, the number of sites at the final layer is therefore $n_{\mathrm{h}} = 2\pi r_+/L$.

This further tells us that the MERA truncates after a finite number of layers m, given by

$$m = \log_k \left(\frac{n_\mathrm{b}}{n_\mathrm{h}} \right) = \log_k \frac{1}{2\pi T a} \,. \tag{4.42}$$

This coincides with the conclusion in Refs. [101,102] that the MERA representation of a thermal state is obtained after $\mathcal{O}(\log_k(1/T))$ iterations of coarse-graining.

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

$$S_\mathrm{MERA}(B) = 4\eta_B f_B(k) \log_k \left(\frac{\ell_0}{\ell_\mathrm{h}} \right) \ln \chi + \ell_\mathrm{h} \ln \chi \,. \tag{4.43}$$

Recall that the entanglement entropy of a single interval B of coordinate length x_0 in a CFT at finite temperature [62] is given, up to a non-universal constant, by

$$S_\mathrm{CFT}(B) = \frac{c}{3} \ln \left(\frac{1}{\pi a T} \sinh \pi x_0 T \right) \,, \tag{4.44}$$

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

When $T \to 0$ in Eq. (4.44), we recover the usual result (4.9). In the $T \to \infty$ limit, the von Neumann entropy gives the usual thermal entropy as entanglement vanishes. Taking $T x_0 \gg 1$, the leading and subleading contributions to the entanglement entropy are

$$S_\mathrm{CFT} = \frac{c}{3} \pi x_0 T + \frac{c}{3} \ln \frac{1}{2\pi a T} \,, \tag{4.45}$$

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

Now let us consider a finite-temperature CFT that is dual to a BTZ black hole with horizon temperature $T = r_+/2\pi L^2$. In terms of geometric MERA parameters, we find that Eq. (4.45) becomes

$$S_{\mathrm{CFT}} = \frac{c}{6}\,\ell_{\mathrm{h}} + \frac{c}{3} m \ln k\,. \tag{4.46}$$

Here we used the fact that $\ell_{\mathrm{h}} = x_0 r_+/L^2$ as well as Eq. (4.42), where we note that m can also be written as $\log_k(\ell_{\mathrm{b}}/\ell_{\mathrm{h}})$. The result (4.46) coincides precisely with the extensive and subextensive contributions calculated using the MERA in Eq. (4.43) provided that $c/\ln\chi \sim \mathcal{O}(1)$. Therefore, we find that the truncated MERA correctly captures the entanglement structure of thermal CFT states and their dual BTZ spacetimes. These conclusions are in agreement with those in Refs. [85, 102].

As a check of the claim that c and $\ln\chi$ should be of the same order, we can compare the horizon entropy given by the contribution from the sites at the final layer with the Bekenstein–Hawking entropy (4.41) of a BTZ black hole. There are n_{h} sites comprising the horizon, each with Hilbert space dimension χ. The system is in the infinite-temperature limit — and hence described by a maximally mixed density matrix, with entropy contribution $\ln\chi$ from each site — so

$$S_{\mathrm{horizon}} = n_{\mathrm{h}} \ln\chi\,. \tag{4.47}$$

Making use of the Brown–Henneaux relation and requiring that the entropy (4.47) coincide with the Beckenstein-Hawking entropy, we again find that $c/\ln\chi \sim \mathcal{O}(1)$. More specifically, taking the counting to be precise, we find that

$$c/\ln\chi = 6\,, \tag{4.48}$$

which is qualitatively in agreement with the previous conclusion (4.20) that the Hilbert space dimension must be exponentially large in c.

With this relation, the extensive terms in Eqs. (4.43) and (4.46) agree precisely. Further identifying the subextensive terms, we find $\eta_B f_B(k) = (\ln k)/2$. If we then

impose the constraint (4.18), we find that

$$\frac{k \ln k}{2(k-1)} \leq 1 \,. \tag{4.49}$$

This last inequality exactly reproduces Eq. (4.35) and thus constrains k to be 2, 3, or 4. Interestingly, we have found the weaker of the two bounds derived in Sec. 4.5, without needing to consider the Bousso bound.

As desired, the truncated MERA computation of entanglement entropy agrees with the expected entanglement entropy given by the application of the Ryu–Takayanagi formula to the length of the minimal surface in a BTZ spacetime. The fact that the results of matching boundary entanglement entropy given in Sec. 4.4 further hold in BTZ spacetimes might not be too surprising given that such spacetimes are quotients of pure AdS_3.

Part II

Cosmology

Chapter 5

DE SITTER SPACE AS A TENSOR NETWORK

We investigate the proposed connection between de Sitter spacetime and the MERA (Multiscale Entanglement Renormalization Ansatz) tensor network, and ask what can be learned via such a construction. We show that the quantum state obeys a cosmic no-hair theorem: the reduced density operator describing a causal patch of the MERA asymptotes to a fixed point of a quantum channel, just as spacetimes with a positive cosmological constant asymptote to de Sitter. The MERA is potentially compatible with a weak form of complementarity (local physics only describes single patches at a time, but the overall Hilbert space is infinite-dimensional) or, with certain specific modifications to the tensor structure, a strong form (the entire theory describes only a single patch plus its horizon, in a finite-dimensional Hilbert space). We also suggest that de Sitter evolution has an interpretation in terms of circuit complexity, as has been conjectured for anti-de Sitter space.

This chapter was published as Ref. [4], N. Bao, C. Cao, S. M. Carroll, and A. Chatwin-Davies, "De Sitter space as a tensor network: Cosmic no-hair, complementarity, and complexity," Phys. Rev. D ***96*** *(2017) 123536,* `arXiv:1709.03513`.

5.1 Introduction

Even in the absence of a completely-formulated theory of quantum gravity, a great deal can be learned by combining insights from classical gravity, semiclassical entropy bounds, the principles of holography and complementarity, and the general structure of quantum mechanics. A natural testing ground for such ideas is de Sitter space, a maximally symmetric spacetime featuring static causal patches with a finite entropy. De Sitter is also of obvious phenomenological relevance, given the positive value of the cosmological constant in the real world. In this paper we apply ideas from quantum circuits and tensor networks to investigate quantum properties of de Sitter

on super-horizon scales.

The Multiscale Entanglement Renormalization Ansatz (MERA) is a well-studied tensor network that was originally developed to find ground states of 1+1 dimensional condensed matter theories [75]. In recent years, an interesting connection has been drawn between the MERA and AdS_3/CFT_2, by way of using the MERA to discretize the AdS space [76, 86]. The argument was made that this could be seen as a way of emerging AdS space from the boundary CFT, thus establishing AdS/CFT as a theory in which bulk spacetime emerges from entanglement properties on the boundary. Further work exploring this direction and generalizing it to other types of tensor networks has been done by [31, 54, 77], and a p-adic approach to AdS/CFT using trees is explored by [103, 104]. However, the AdS/MERA correspondence seems to have tensions with other known results in holography. For example, it is puzzling that AdS/MERA appears to suggest a "bulk geometry" in the form of a tensor network even for a CFT with a small central charge. Additionally, it needs to satisfy a set of stringent constraints, brought on by the fact that it is supposed to duplicate the established results of AdS/CFT [15, 17]. It appears that AdS/MERA in its simplest form is not able to satisfy all of the constraints imposed by holography with AdS geometry [3, 105, 106], although extensions may be able circumvent this difficulty [107].

There is also considerable interest in studying a more general notion of geometry from entanglement beyond the context of AdS/CFT [20], where geometries are related to our physical universe [108, 109]. A connection between the MERA and de Sitter spacetime has been suggested, where we think of the tensors as describing time evolution, rather than as relating different spatial regions [79, 105, 110]. In the case of 1+1 dimensions, it is also claimed [105] that the MERA can be thought of as a discretization of a slice in the "kinematic space" [78, 111], which corresponds to the space of geodesics in the hyperbolic plane in the particular case of AdS_3/CFT_2. This beautifully illustrates a correspondence between regions in the dual kinematic space, which take on information-theoretic interpretations, and the individual tensors

localized in the MERA. More tentatively, quantum circuits have been proposed as a way of studying realistic cosmological evolution from inflation to the present epoch and beyond [112].

In this paper we investigate this proposed connection between the MERA and de Sitter, under the assumption that a MERA-like circuit is able to simulate effective quantum gravitational dynamics on super-Hubble scales for some subset of quantum states in a theory of quantum gravity. We show that the structure of the MERA is able to reproduce some desirable features of evolution in a de Sitter background. In particular, we identify a scale invariant past causal cone as the static patch where an analogous light-like surface functions as the cosmic horizon. Then we show that a version of the cosmic no-hair theorem can be derived from the fixed point of the quantum channel, whereby any state will asymptote to the channel fixed point at future infinity. We next examine the issue of horizon complementarity in the MERA context, and argue that the global and local descriptions of de Sitter [113, 114] can be equivalent up to a unitary change of basis. We observe similarities between a strong version of local de Sitter and the implementation of a quantum error correcting code. Lastly, we derive a bound on the quantum complexity of the MERA circuit, and show that the complexity scales in a manner that is consistent with the "complexity equals action" conjecture [115].

5.2 The MERA and the de Sitter causal patch

In Fig. 5.1 we illustrate the MERA tensor network. In its original conception as an ansatz for constructing ground states of 1-d spin systems, one starts with a simple quantum state at the top of the diagram, and propagates it downward through a series of gates to a final state at the bottom. Each line represents a factor of Hilbert space, which might be quite high-dimensional. Moving downward is the "fine-graining" direction, and upward is "coarse-graining." The square gates are "disentanglers" (although they create entanglement as we flow downward), which take two factors in and output another two factors. The triangular gates are "isometries,"

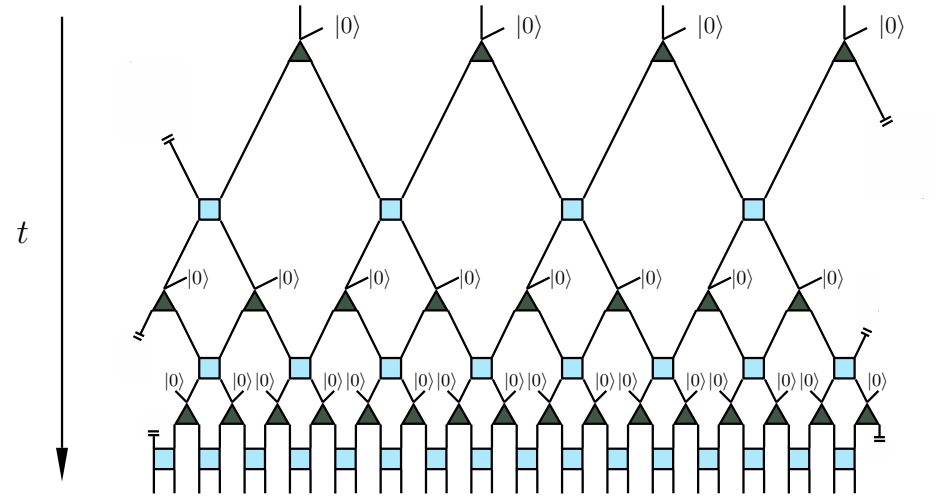

Figure 5.1: A periodic binary MERA. The green triangles denote the isometries and the blue squares denote the disentanglers. The kets labeled $|0\rangle$ are ancilla states inserted into each isometry. The action of the circuit is to take a state at the top and evolve it downward. In anticipation of the connection to de Sitter, the fine-graining direction is labelled as the direction of increasing t.

which can be thought of as taking in a single factor and outputting two factors; alternatively, we can imagine inputting two factors, one of which is a fixed state $|0\rangle$, and outputting another two, so that the total dimensionality entering and exiting each tensor is equal. We will adopt the latter perspective in this paper. It is often convenient to consider generalizations where $k > 2$ factors enter and exit each tensor.

In the AdS/MERA correspondence, tensors are taken to represent factors of Hilbert space, and the two-dimensional geometry of the graph is mapped to the hyperbolic plane. Here, where we are interested in studying a dS/MERA correspondence, flow through the circuit represents evolution through time. Note that, while it is common in general relativity to draw spacetime diagrams with the future at the top, the convention in quantum circuits for MERA is to start with one or more "top tensors" and evolve downward. Here we will stick to the conventions of the respective communities; time flows downward in MERA circuit diagrams, and upward in spacetime diagrams.[1]

In this work, we will mostly be concerned with MERAs that are scale and transla-

[1] We will occasionally draw circuit diagrams in which time flows from left to right, just to keep things lively.

tionally invariant (the same disentanglers and isometries appear everywhere in the network). We use the term "site" in the MERA to refer to a Hilbert space factor that lives on a leg that exits a disentangler (or equivalently, that enters an isometry). When the MERA is used as a variational ansatz for a physical system like a spin chain, the collection of sites at any given layer corresponds to the state of the physical lattice at that renormalization scale. For more extensive reviews of tensor networks and the MERA see [3, 80, 100].

Viewed as a circuit in which the fine-graining direction corresponds to the future or past direction (away from the de Sitter throat), the MERA reproduces the causal structure of de Sitter spacetime [79, 105, 110]. Recently, as a part of their studies of kinematic space, Czech *et al.* further pointed out that there is a natural way of associating the MERA with half of the 1+1-dimensional de Sitter manifold [105]. Here we briefly explain how this works.

Let $\mathcal{M}$ be 1+1-dimensional de Sitter spacetime with the usual global coordinatization:

$$ds^2 = \ell_{\mathrm{dS}}^2(-dt^2 + \cosh^2 t\, d\theta^2). \tag{5.1}$$

The timelike coordinate t takes all real values, and θ is an angular coordinate that is 2π-periodic. In these coordinates, $\mathcal{M}$ looks like a hyperboloid whose constant-t sections are circles that attain a minimum radius at $t = 0$ and that grow in either

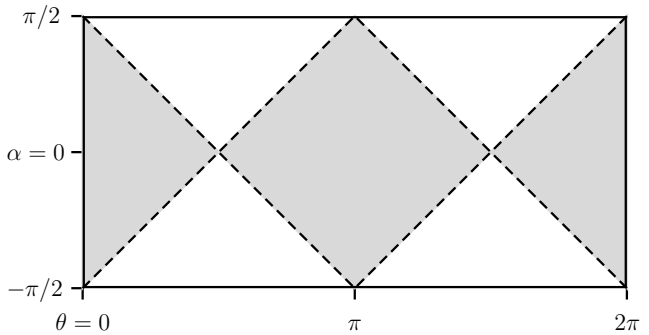

Figure 5.2: The Penrose diagram of global (1+1)-dimensional de Sitter spacetime. As this is a spacetime diagram, time now runs from bottom to top. The boundaries of two complete disjoint causal patches, one centered at $\theta = 0$ and the other centered at $\theta = \pi$, are drawn with a dashed line, and the interiors of the patches are shaded. Light rays travel along 45° lines in this diagram.

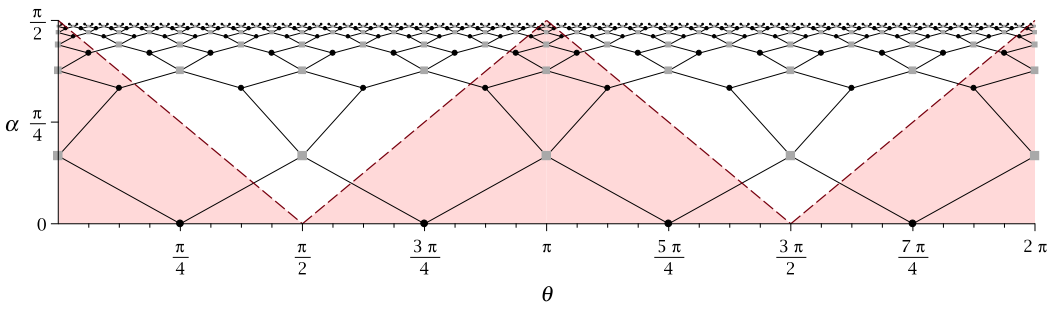

Figure 5.3: A geometric de Sitter-MERA correspondence, mapping the MERA circuit to the top half of the de Sitter geometry. Note that the fine-graining direction of the MERA in this diagram points upward to match the future direction in the Penrose diagram. The domain of dependence of any pair of adjacent sites in the initial layer of the MERA is entirely contained within a single static patch in de Sitter. Two of the four possible static patch interiors are shaded in red. (The other two static patches are centered at $\theta = \pi/2$ and $\theta = 3\pi/2$.)

direction away from $t = 0$. The proper radius at $t = 0$ is equal to ℓ_{dS}, which is called the de Sitter radius. A convenient coordinate transformation is to set $\cosh t = \sec \alpha$, under which the metric becomes conformally flat:

$$ds^2 = \frac{\ell_{\text{dS}}^2}{\cos^2 \alpha} \left(-d\alpha^2 + d\theta^2\right). \tag{5.2}$$

Because of this, the full de Sitter manifold is often represented by a rectangle in the θ-α plane with $-\pi/2 < \alpha < \pi/2$ and $0 < \theta < 2\pi$, as in the Penrose diagram of de Sitter, Fig. 5.2.

Consider now the top half of the de Sitter manifold with $t \geq 0$ (or $0 \leq \alpha < \pi/2$). Starting at $t_0 \equiv 0$, the length of the constant-t_n slice doubles at every subsequent time $t_n = \operatorname{arccosh} 2^n$ with $n = 1, 2, \ldots$ This suggests identifying the top of a translationally invariant binary MERA with the $t_0 = 0$ slice, and subsequent layers of the MERA with the subsequent t_n slices, so that the MERA describes the top half of the de Sitter hyperboloid. This identification is illustrated in Fig. 5.3, in which the sites of the n^{th} layer of the MERA have been chosen to lie at the angles

$$\theta_j^{(n)} = \frac{\pi}{2^{n+1}} \left(j + \frac{1}{2}\right) \qquad j = 0, \ldots, 2^{n+2} - 1\,. \tag{5.3}$$

The fact that the top of the MERA was chosen to have four sites was no coincidence. With this choice, the *future domain of dependence* of any two adjacent sites at the

top of the MERA precisely coincides with (the top half of) a single static patch of de Sitter. Or, to use terminology that is more familiar in the MERA literature, each static patch of de Sitter that is centered at $\theta = 0$, $\pi/2$, π, or $3\pi/2$ coincides with a causal cone [87] in the MERA such that every layer of the causal cone contains precisely two sites of the MERA (i.e., the causal cone is stationary).

Let us elaborate a bit on the terminology above. First, recall how a domain of dependence is defined on a smooth manifold:

Definition 5.2.1. *Let $S \subset \mathcal{M}$ be a subset of a smooth Lorentzian manifold $\mathcal{M}$. The future (resp. past) domain of dependence of S is the set of all points $p \in \mathcal{M}$ such that every past (resp. future) inextensible causal curve through p intersects S.*

This suggests the following analogous definition for a domain of dependence in a MERA:

Definition 5.2.2. *Let S be a collection of sites in a MERA. The future (resp. past) domain of dependence of S is the set of all MERA sites p such that starting at p and moving only in the past, or coarse-graining direction (resp future, or fine-graining direction), one inevitably arrives at a site in S.*

In de Sitter space, the proper radius of the cosmological horizon is constant. Given an inextendible timelike geodesic, a static patch is defined as the set of all points connected to that geodesic by both past- and future-oriented causal curves, and its size is given by the horizon radius. In particular, in 1+1 dimensions the horizon radius is $\pi\ell_{\mathrm{dS}}/2$. Within a constant-t slice, a horizon volume is an interval of proper length $\pi\ell_{\mathrm{dS}}$, and static patches are diamonds in the Penrose diagram (cf. Fig. 5.2).

In line with [105], we here adopt a correspondence between the MERA and half of the full 1+1-dimensional de Sitter manifold in which stationary causal cones in the MERA are in correspondence with static patches of de Sitter. In the spirit of tensor-network/spacetime correspondences, one should think of the MERA and the state that it describes as some state of quantum gravity describing quantum fields

evolving in a semiclassical de Sitter background. In other words, despite lacking an explicit theory of quantum gravity, we suggest that some aspects of the effective dynamics for a quantum gravity state that describes classical de Sitter spacetime can be described and organized at a fundamental level by a suitably-chosen MERA.

In this picture, each site of the MERA carries a Hilbert space $\mathcal{H}_*$, and the Hilbert space that corresponds to a given horizon volume (call it $\mathcal{H}_{\text{static}}$) is the tensor product of the Hilbert spaces of the sites that lie within the horizon. We do not count the Hilbert spaces that correspond to unentangled ancillae as part of the static patch Hilbert space, since we only attach a spacetime interpretation to entangled degrees of freedom in the MERA proper. To be consistent with the Gibbons-Hawking entropy of de Sitter spacetime [116], it should be that $\ln \dim \mathcal{H}_{\text{static}} \sim S_{\text{dS}}$, where S_{dS} is the de Sitter entropy. Therefore, for our Universe, where $S_{\text{dS}} \sim 10^{122}$, the corresponding bond dimension (i.e., the dimensionality of $\mathcal{H}_*$) is of order $\dim \mathcal{H}_* \sim \exp(10^{122})$ per site.

This is a very coarse-grained description of de Sitter spacetime. For a binary MERA, there are only two sites per horizon volume, and layers of the MERA within a static patch are separated by cosmological timescales. Furthermore, a binary MERA only accommodates 4 distinct static patches (Fig. 5.2). We imagine, however, that it should be possible to refine this horizon-scale description via, e.g., local gadget expansions, in which the large Hilbert space $\mathcal{H}_*$ could be factorized according to sub-horizon locality. This perhaps can be achieved by some version of cMERA [88, 89, 117].

One might wonder whether it is possible to pack more MERA sites into a single slice of the static patch by starting with more sites at the top of the MERA, or by considering a MERA with a larger branching factor. The number of sites at the top of the MERA is fixed by the number of sites per layer in the stationary causal cone, however. If the stationary causal cone has m sites per layer, then the $t = 0$ slice contains $2m$ sites. The reason is simply because the $t = 0$ slice of de Sitter contains exactly two disjoint horizon volumes. The quantity m in turn is fixed by the

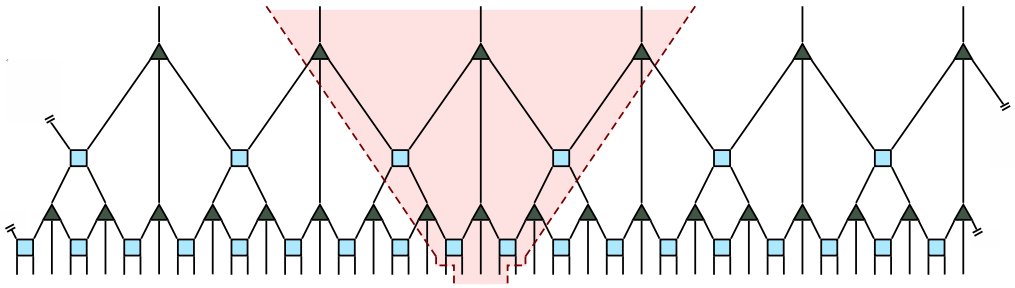

Figure 5.4: A ternary MERA. Ancillae are suppressed in this diagram. A stationary causal cone with three sites per layer is indicated by the shaded region.

branching factor and the structure of the MERA. For a binary MERA, a stationary causal cone always has $m = 2$ sites per layer. A ternary MERA has $m = 3$ sites per layer in a stationary causal cone (Fig. 5.4). However, in general for a k-nary MERA, in which the number of sites increases k-fold in each layer of the MERA, there can only ever be $m = 2$ or 3 sites per layer in a stationary causal cone. Further details of stationary causal cones and a proof of this last fact are given in Sec. 5.7.

Unfortunately, the global de Sitter-MERA correspondence as formulated on a (hyper)cubic lattice does not easily generalize to higher dimensions due to discretization artifacts. The possibility of a de Sitter-MERA correspondence in higher dimensions is discussed in Sec. 5.8.

5.3 Cosmic No-Hair as a channel property

Via the correspondence described above, each constant-t slice of a de Sitter static patch is assigned a Hilbert space

$$\mathcal{H}_{\text{static}} = \mathcal{H}_* \otimes \mathcal{H}_*, \tag{5.4}$$

where $\mathcal{H}_*$ is the Hilbert space of a single MERA site. If we restrict our attention to a single static patch, then the MERA also defines a superoperator, $\mathcal{E}$, which maps a state in $\mathcal{H}_{\text{static}}$ forward by one Hubble time to a state on the next slice. With the disentanglers and isometries held fixed and uniform across the MERA, the action of $\mathcal{E}$ may be written explicitly as

$$\mathcal{E}(\rho) = U_{BC}\ \mathrm{Tr}_{AD}\left[V_{AB} \otimes V_{CD}(|0\rangle\langle 0|_A \otimes \rho_{BC} \otimes |0\rangle\langle 0|_D)V_{AB}^\dagger \otimes V_{CD}^\dagger\right] U_{BC}^\dagger. \tag{5.5}$$

The labels A, B, C, and D indicate on which Hilbert space factors operators act, but we may subsequently omit them when it does not cause confusion. The ancillae are labelled by A and D, and $\mathcal{H}_{\text{static}}$ is labelled by B and C, cf. Fig. 5.5.

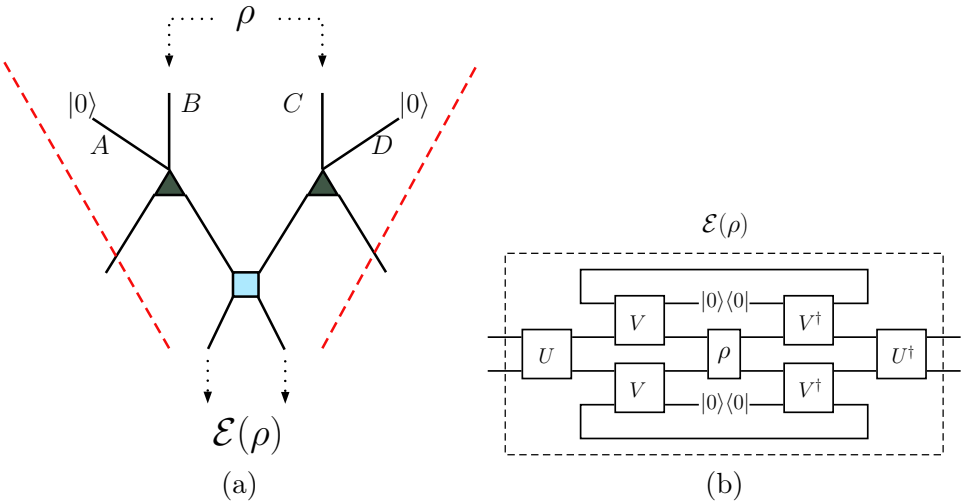

Figure 5.5: (a) A single step of the MERA within the causal patch, viewed as a channel $\mathcal{E}$, and (b) the equivalent circuit diagram. Time runs in the downward direction in (a).

In the MERA literature, $\mathcal{E}$ is known as the *descending superoperator* [80,118]. It is a quantum channel by construction, i.e., it is completely positive and trace-preserving on the set of states (density operators), which for future reference we will denote by $\mathcal{S}(\mathcal{H}_{\text{static}})$. In precise language, given a Hilbert space $\mathcal{H}$, if $\mathscr{H}(\mathcal{H})$ denotes the space of Hermitian operators on $\mathcal{H}$, then the set of states is

$$\mathcal{S}(\mathcal{H}) \equiv \{\rho \in \mathscr{H}(\mathcal{H}) \mid \text{Tr } \rho = 1,\ \langle\psi|\rho|\psi\rangle \geq 0 \ \ \forall\ |\psi\rangle \in \mathcal{H}\}\,. \tag{5.6}$$

Consider now starting at some given layer with a state $\rho_0 \in \mathcal{S}(\mathcal{H}_{\text{static}})$ and repeatedly applying the map $\mathcal{E}$. Intuitively, every application of $\mathcal{E}$ dilutes the original state ρ_0 by entangling it with the same ancillary state $|00\rangle\langle 00|_{AD}$ before taking a partial trace, at which point information about ρ_0 flows out of the static patch. It is therefore natural (and correct) to expect that the state on the static patch should settle down to a future asymptotic steady state, regardless of the initial state ρ_0.

We will make this expectation rigorous below, but first we note that this observation suggests a sort of *cosmic no-hair theorem* for the de Sitter-MERA correspondence.

In classical general relativity, a cosmic no-hair theorem is roughly the statement that a positive cosmological constant causes a spacetime to asymptotically tend to a de Sitter state in the future. The following theorem of Wald pertaining to Bianchi spacetimes, which are homogeneous but anisotropic cosmological models, is perhaps the most precise statement of a cosmic no-hair theorem [119]:

Theorem 5.3.1 (Wald)**.** *All Bianchi spacetimes (except certain strongly-curved Bianchi IX spacetimes) that are initially expanding, that have a positive cosmological constant, and whose matter content obeys the strong and dominant energy conditions asymptote to de Sitter in the future.*

Various generalizations and variations of this theorem exist in the literature [5, 120–128]. In particular, quantum cosmic no-hair theorems show that the quantum states of fields tend to their respective vacuum states on an asymptotically de Sitter background [129–131]. The MERA results here are reminiscent of these quantum cosmic no-hair theorems.

Let us now add some rigor to the above observations. When $\mathcal{H}$ is finite-dimensional, quantum channels are necessarily contractions on $\mathcal{S}(\mathcal{H})$ [132]. Recall that a linear map $T : X \to X$ on a Banach space X is a *contraction* if there exists $0 < \kappa \leq 1$ such that $d(T(x_1), T(x_2)) \leq \kappa\, d(x_1, x_2)$ for all $x_1, x_2 \in X$, where d is the metric on X. For $\mathcal{S}(\mathcal{H})$, the metric is most commonly defined using the 1-norm,

$$d(\rho, \sigma) \equiv ||\rho - \sigma||_1, \tag{5.7}$$

where $||A||_1 = \mathrm{Tr}\sqrt{A^\dagger A}$ for any linear operator A.[2] A contraction is *strict* when $0 < \kappa < 1$, in which case the contraction mapping principle guarantees that there is a unique fixed point $x_\star \in X$ such that $T(x_\star) = x_\star$. Furthermore, the sequence $\{T^n(x_0)\}_{n=1}^\infty$ converges to the fixed point $x_\star$ for any choice of the starting point x_0.

Quantum channels need not be strict contractions in general; however, it is certainly easy to write down channels that are strict contractions [132]. Returning to the

[2] All norms are equivalent in finite dimensions, i.e., for any two norms $\|\cdot\|_a$ and $\|\cdot\|_b$, there exist constants $m > 0$ and $M > 0$ such that $m\|v\|_a \leq \|v\|_b \leq M\|v\|_a$ for all v in the normed space.

de Sitter-MERA correspondence, we may simply suppose that the disentanglers U and isometries V are chosen such that the superoperator $\mathcal{E}$ is a strict contraction. Moreover, numerical assays seem to indicate that this is generally the case for random U and V [80,118]. Our intuition that the state in a causal patch should tend to some asymptotic fixed state in the future is therefore warranted.

Regardless of the channel's contractive properties, it is easy to see that $\mathcal{E}$ has at least one fixed point by examining its adjoint. To define the adjoint, take the domain of $\mathcal{E}$ to be the space of Hermitian operators, $\mathscr{H}(\mathcal{H})$, which is closed under addition and multiplication by real numbers. The space $\mathscr{H}(\mathcal{H})$ with the Frobenius inner product

$$\langle T, S\rangle \equiv \mathrm{Tr}\left(S^\dagger T\right) \tag{5.8}$$

is then a Hilbert space over the real numbers. As usual, the adjoint operator is defined by the relation $\langle \mathcal{E}(T), S\rangle = \langle T, \mathcal{E}^\dagger(S)\rangle$. Using this definition, it is straightforward to show that the action of $\mathcal{E}^\dagger$ is

$$\mathcal{E}^\dagger(S) = {}_{AD}\langle 00|V_{AB}^\dagger V_{CD}^\dagger \left[I_{AD}\otimes (U^\dagger S U)_{BC}\right] V_{AB}V_{CD}|00\rangle_{AD}\,. \tag{5.9}$$

In the MERA literature, $\mathcal{E}^\dagger$ is known as the *ascending superoperator*. In this form, it is clear that the identity operator is an eigenvector of $\mathcal{E}^\dagger$ with eigenvalue $\lambda = 1$. Therefore, $\bar{\lambda} = \lambda = 1$ is also in the spectrum of $\mathcal{E}$, or in other words, $\mathcal{E}$ necessarily has a fixed point.

That $\lambda = 1$ is an eigenvalue of $\mathcal{E}$ is well-known [80, 118]; however, we exhibited $\mathcal{E}^\dagger$ because it clearly shows that, in general, $\mathcal{E}$ is not self-adjoint. In particular, this means that the eigenvector of $\mathcal{E}$ to the eigenvalue 1, call it $\rho_\star$, is not trivially the identity operator. An interesting question is how much freedom is possible in choosing $\rho_\star$ by specifying the disentanglers and isometries U and V. Clearly there are families fixed points. For example, if $\rho_\star$ is such that $\mathcal{E}(\rho_\star) = \rho_\star$ for a given choice of U and V, then $\tilde{\rho}_\star \equiv (W^\dagger \otimes W^\dagger)\rho_\star(W\otimes W)$ is the fixed point of the channel $\tilde{\mathcal{E}}$ with $\tilde{U} = W^\dagger U$ and $\tilde{V} = (I\otimes W)V$ for any unitary operator W on $\mathcal{H}_*$. From exactly what subset of $\mathcal{S}(\mathcal{H})$ the fixed point $\rho_\star$ may be chosen is an open problem.

5.4 Global de Sitter and Complementarity

In classical general relativity, there are no barriers to describing de Sitter spacetime in a global way. However, in light of complementarity [23], an interesting question is whether quantum gravity also accommodates a global description of de Sitter, or whether a fully quantum theory only exists on a single causal patch. We will suggest that a local picture (describing only a single patch) is possible via the MERA if the Hamiltonian is essentially time-dependent; as a result, this perspective also avoids Poincaré recurrences.

Complementarity, as it was originally envisioned for black holes, asserts that the ability of an observer to describe the region around them in terms of local quantum field theory on a smooth spacetime background does not extend into the unobservable region behind a horizon. For example, when describing physics outside of the black hole in a black hole spacetime, one should think of all of the black hole's degrees of freedom as residing just above its apparent horizon on a stretched horizon [133]. Nevertheless (and neglecting possible issues regarding firewalls [18]), there should also exist a complementary description of the black hole that is appropriate to, e.g., an observer who crosses the horizon, where the black hole interior is very much a real place. Any possible discrepancies in these two descriptions are then purportedly resolved by the fact that an observer who crosses the horizon becomes causally disconnected from the black hole exterior, and so information about these discrepancies cannot be communicated to the exterior. Applied to de Sitter cosmology, horizon complementarity suggests that a single observer can only describe physics using local quantum field theory in a region that stretches out to the horizon, but no farther. To this observer, the only sign of the rest of the universe is encoded on a stretched horizon. If one considers two observers that have overlapping horizon volumes, then there is presumably some partial mapping between their respective local descriptions of physics.

The question then arises as to whether an infinitely big spacetime outside the de Sitter horizon actually exists in this picture. A weak version of complementarity

might posit that it does, but that its existence cannot be described by any one observer; the underlying quantum theory would nevertheless still describe states in an infinite-dimensional Hilbert space. A stronger version would postulate that the entire quantum theory has a finite-dimensional Hilbert space (with dimension of order $e^{S_{\rm dS}}$), and all that exists can be described by a single Hubble patch and its horizon [113, 114, 134–140]. The descriptions of physics in different horizon volumes contained in different causal patches are then related by a global unitary transformation. The distinction might seem academic, but is actually crucial: unitary evolution with a time-independent Hamiltonian in a finite-dimensional Hilbert space leads to Poincaré recurrences and Boltzmann brains [139, 141, 142], which can be avoided if Hilbert space is infinite-dimensional [143].

Let us refer to the weak complementarity perspective as the "global" view (different regions of the classical de Sitter spacetime have an independent existence, and Hilbert space is infinite-dimensional), and the strong complementarity perspective as the "local" view (there is only one patch worth of information, and Hilbert space is finite-dimensional). The MERA tensor network, we will argue, can accommodate the local description, and with a bit of modification, the global description as well. We find that there is a natural sense in which the information associated with any single static patch can be localized on the static patch and its horizon. We then propose a modified network that we call SCMERA ("Strong Complementarity MERA") that could, in principle, capture the local strong complementarity view. In order to have consistent time-evolution in the SCMERA, we will see that it is effectively generated by a time-dependent Hamiltonian, i.e., the unitary operator that maps a layer in the SCMERA to the next layer changes as a function of depth in the network. While such evolution is in tension with our expectations in cosmology, where the Hamiltonian evolution should be time-independent, it does avoid certain undesirable phenomena like Poincaré recurrences. Given how little we know about quantum cosmology, it seems worth keeping different perspectives in mind.

5.4.1 Slicing, weak complementarity, and pseudo-holography

A notable feature of the MERA is that it naturally provides a way to both define different Cauchy slices and relate the states defined on them. Up until now, we have thought of states in global de Sitter as being defined on constant time slices, or in other words, on a single layer at constant depth in the MERA. However, given such a state that we label by $|\Psi\rangle_{\text{dS}}$, by picking some collection of sites on which it is defined, one can define a new state $|\tilde{\Psi}\rangle_{\text{dS}}$ (which is in a tensor product with some collection of n ancillae) and a new Cauchy slice by pushing the state on the chosen sites back up (i.e., backwards in time) through the MERA. In other words, $|\Psi\rangle_{\text{dS}}$ and $|\tilde{\Psi}\rangle_{\text{dS}} \otimes |0\rangle^{\otimes n}$ are related by partial unitary evolution, and the horizontal cut through the MERA on which $|\tilde{\Psi}\rangle_{\text{dS}} \otimes |0\rangle^{\otimes n}$ is defined constitutes a new Cauchy slice. In particular, given a static patch, the state $|\Psi\rangle_{\text{dS}}$ can be pushed back up through the MERA in this way so that the resulting state is supported entirely on the sites that comprise $\mathcal{H}_{\text{static}}$ and sites that are on the lightlike horizon, as illustrated in Fig. 5.6. Note that this wouldn't be possible for a generic state living on a constant $t = T$ slice in the Hilbert space of the complete theory, but can be done for the specific states that arise via the MERA from the initial state at $t = 0$ (the top tensors).

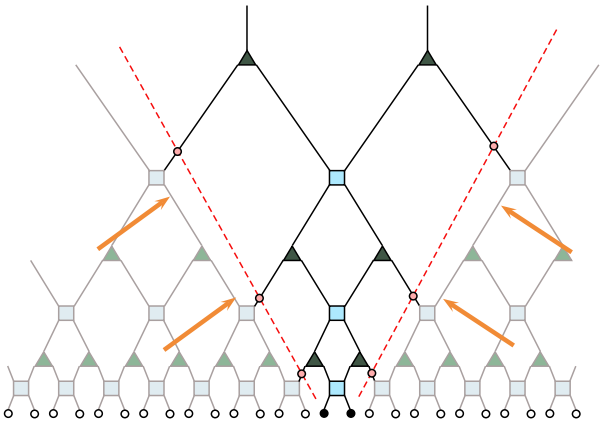

Figure 5.6: Sites outside the horizon at any given layer (indicated by white dots) are unitarily related, via the MERA, to a state on the horizon (indicated by red dots) and a collection of ancillae (not shown), $|\tilde{\Psi}\rangle_{\text{dS}} \otimes |0\rangle^{\otimes n}$. A state $|\Psi\rangle_{\text{dS}}$ corresponding to the de Sitter spatial slice is prepared at the bottom layer. The sites inside the static patch are indicated by the filled black dots.

The observation above suggests a toy model for weak complementarity as well as a sort of "pseudo-holography." The network clearly admits a global de Sitter description on constant time slices, but a more observer-centric view of the local patch consists of the state defined on $\mathcal{H}_{\text{static}}$ and a collection of horizon sites, as discussed above and shown in Figure 5.6. For a stationary observer $\mathcal{O}_A$ who travels along a timelike geodesic at the center of the static patch, all information relevant to $\mathcal{O}_A$'s local description of physics is given by the degrees of freedom in the static patch interior. The information about the exterior is encoded in the degrees of freedom that reside on the horizon. However, for another observer $\mathcal{O}_B$ who travels away from $\mathcal{O}_A$ and leaves the patch, their surrounding spacetime geometry and description of the quantum state can be "manufactured" by propagating $\mathcal{O}_A$'s horizon degrees of freedom down through the MERA. In this way, the region that is accessible to $\mathcal{O}_B$ is realized by decompressing [105] the information that is contained on $\mathcal{O}_A$'s horizon. The information that was previously understood to have localized on the horizon for $\mathcal{O}_A$ is, up to inclusion of ancillae, unitarily transformed to a state defined on spacetime that is to the exterior of $\mathcal{O}_A$'s static patch. This map between the local descriptions of different observers is a realization of weak complementarity, with information about spacetime to the exterior of an observer's cosmic horizon being encoded on the horizon in a way that seems holographic.

This picture of weak complementarity is not really holographic, however, because the number of apparent degrees of freedom associated with the horizon increases toward the future in the MERA, i.e., the number of horizon sites grows with every subsequent layer. In a true holographic model, the size of the boundary Hilbert space should remain constant. We investigate this possibility, or in other words, the possibility of strong complementarity, in the next section.

5.4.2 Strong Complementarity, recoverability, and quantum error correction

In the local, strong complementarity picture, the degrees of freedom represented by the static patch of a single observer, plus those on the corresponding horizon, to-

gether describe a closed system constituting the entirety of Hilbert space, which is correspondingly finite-dimensional. Ordinarily, assuming a time-independent Hamiltonian, such a setup would lead to recurrences and Boltzmann brains. What we will find, however, is that it is more natural from the MERA perspective to imagine evolution inside the patch that is equivalent to a time-dependent Hamiltonian. (Cosmological evolution with a time-dependent Hamiltonian also plays a role in Banks and Fischler's approach to holographic spacetime [144–146].)

A local picture is possible in the MERA because of its particular circuit construction that begins with a finite number of inputs (4 for a binary MERA), where only two non-overlapping static patches at $t = 0$ are present. Consequently, the total number of quantum degrees of freedom for the input is limited to that of two non-overlapping patches and is, of course, finite. Let $\chi_* \equiv \dim \mathcal{H}_*$ denote the dimension of the Hilbert space of a single MERA site (the bond dimension). Then, even though the number of sites in the MERA grows as a function of depth, the global state at any given subsequent layer of the MERA only resides in a subspace of dimension χ_*^4. Because $\dim \mathcal{H}_{\text{static}} = \chi_*^2$ remains the same at every step in the MERA within the static patch, there always exists a purification of the state $\rho_{\text{static}} \in \mathcal{S}(\mathcal{H}_{\text{static}})$ in a Hilbert space with dimension χ_*^2. Therefore, simply by counting Hilbert space dimensions, we could imagine that such a purifying Hilbert space, call it $\mathcal{H}_{\text{horizon}}$, resides on the horizon of the static patch. The horizon state would have to be unitarily related to the global state of the MERA outside the static patch (which is a preferred purification of ρ_{static}).

To turn the network into a description of a single-patch universe, we propose modifying the MERA circuit as follows. First, choose any single static patch in the MERA (cf. Fig. 5.3). At $t = 0$, we identify the degrees of freedom inside a static patch as interior degrees of freedom living in the Hilbert space $\mathcal{H}_{\text{static}}$. The remaining exterior degrees of freedom in the other patch can now be identified with the horizon within the Hilbert space $\mathcal{H}_{\text{horizon}}$, with $\dim \mathcal{H}_{\text{static}} = \dim \mathcal{H}_{\text{horizon}} < \infty$. For a local picture, we preserve the circuit structure for the static patch interior, but now we introduce

separate circuit dynamics for $\mathcal{H}_{\text{horizon}}$, as shown in Fig. 5.7. In particular, a recovery tensor (indicated by the ellipse) acts to extract ancilla states at the horizon. Because the interior network is unchanged, the previous cosmic no-hair result about the interior state continues to hold.

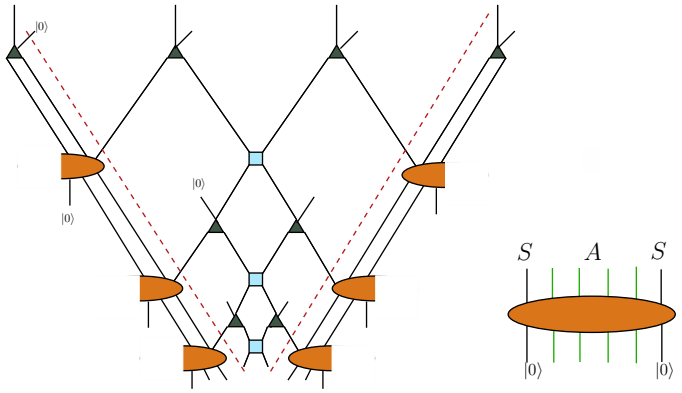

Figure 5.7: The strong complementarian version of the MERA that describes a static patch for a local observer with horizon degrees of freedom. The future direction points downward in the fine-graining direction. Dashed red lines demarcate the interior of the de Sitter static patch. The combined system, including a constant number of ancillae, evolves unitarily. The horizon degrees of freedom at each time step are acted upon by a single recovery tensor (orange ellipse), which serves as a map that distills the same ancillary state (represented by $|0\rangle$ in the figure) that is entangled in the interior at the horizon. (Half-ellipses on opposite sides of the tensor network are identified.) The ancillary system is denoted by S while the horizon degrees of freedom are denoted by A.

This circuit structure constrains the action of the recovery tensor that acts on $\mathcal{H}_{\text{horizon}}$ in Fig. 5.7 if we demand unitary evolution. At each time step, new ancillae are mixed with the interior via the action of the isometries (triangular tensors), and then some information will flow to the horizon and become inaccessible to any interior observer via the action of the disentanglers (square tensors). To be consistent with the literature, label the Hilbert space of the ancillae by S, the static patch Hilbert space by E (i.e., $\mathcal{H}_{\text{static}} \equiv E$), and the horizon Hilbert space by A (i.e., $\mathcal{H}_{\text{horizon}} \equiv A$). If it is always the same ancillary state σ_S (which we have simply taken to be $\sigma_S = |0\rangle\langle 0|_S$ throughout) that gets mixed in via the isometries, then in order to have consistent unitary evolution, it must be that the recovery tensor, which acts on AS, must spit out a state of the form $\rho''_A \otimes \sigma_S$. Put another way, if

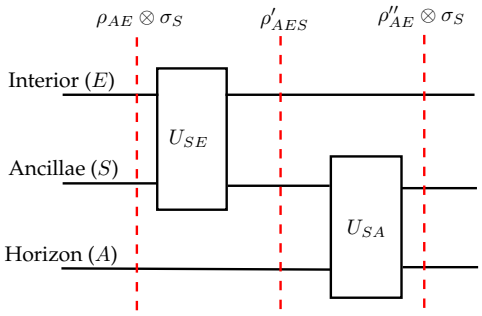

Figure 5.8: Each time step of SCMERA can be condensed into a circuit diagram. The dashed lines mark the resulting quantum state at the end of a subprocess. In the case where the MERA global state is pure, which is the case we consider here, it follows that ρ_{SEA}, ρ'_{SEA}, and ρ''_{SEA} are all pure states.

at every time step we re-introduce a fresh "copy" of the ancillary state σ_S, then unitarity in each time step demands that σ_S be restored after evolving forward in time. (Alternatively we could drop the requirement of unitary evolution; we will return to this possibility at the end of this section.) We call such a circuit for the local picture the Strong Complementarity MERA (SCMERA). The usual global picture can be easily restored by allowing ourselves more ancillary degrees of freedom and replacing the horizon tensors with the usual MERA circuit. As a result, the local and global pictures are related by some global unitary transformation that act on the extended set of ancillae.

Let us ask whether it is possible to have a circuit with the tensor structure in Fig. 5.7 that spits out the state σ_S at every time step. To answer this question, it is useful to analyze the SCMERA circuit from the perspective of recovery maps. At each time step of SCMERA, we can describe the quantum process by

$$\rho_{AES} = \rho_{AE} \otimes \sigma_S \xrightarrow{U_{SE} \otimes I_A} \rho'_{AES} \xrightarrow{U_{SA} \otimes I_E} \rho''_{AES} = \rho''_{AE} \otimes \sigma_S, \tag{5.10}$$

as shown in the quantum circuit diagram in Fig. 5.8. U_{SE} corresponds to the isometries that entangle the ancillae and the interior degrees of freedom, as well as the disentanglers, while U_{SA} acts on the horizon. Since U_{SA}, which corresponds to the elliptical orange tensor in Figure 5.7, must recover the state σ_S, we call it the recovery tensor.

Although the existence of such a recovery tensor is not always guaranteed, we can examine the necessary conditions that these tensors and states must satisfy to allow such a recovery operation. For instance, if the ancillary qudit is always initialized in a fixed vector, e.g., $|0\rangle_S$, or more generally is always chosen from some fixed subspace of S, then one can derive necessary conditions for recoverability by appealing to results from quantum error correction.

To understand the recoverability of the ancillary state, we first consider the action of U_{SE} as a quantum channel on S, $\mathcal{N}_{\rho_{SE}} : S(\mathcal{H}_S) \to S(\mathcal{H}_S)$. This is always possible because the initial state is uncorrelated across S and E. However, because the state in E is in principle arbitrary (and certainly will change at each time step if the mapping is not at a fixed point), the channel can depend on the input ρ_{SE}. Likewise, the recovery tensor will not remain fixed at every time step. This is what we mean when we say that the SCMERA describes evolution that is generated by a time-dependent Hamiltonian; the recovery tensor will change at every time step if it must recover σ_S exactly.

Given such a channel and knowledge of the fixed ancillary state σ, there always exists a process in the reduced system S that recovers σ. Let $\sigma \geq 0$ be the known state in which the ancillary system is initialized. In general, there exists a completely positive trace preserving (CPTP) recovery map R such that $R \circ \mathcal{N}_{\rho_{SE}}(\sigma') = \sigma'$ for all σ' if and only if the monotonicity condition is saturated [147–150]:

$$D(\sigma' \| \sigma) = D(\mathcal{N}_{\rho_{SE}}(\sigma') \| \mathcal{N}_{\rho_{SE}}(\sigma)), \tag{5.11}$$

where $D(\sigma' \| \sigma)$ is the relative entropy between σ' and σ. In particular, σ is always recoverable because the monotonicity condition is trivially saturated when $\sigma' = \sigma$. For the finite-dimensional case, one can construct an explicit Petz recovery map P that will always recover σ':

$$P_{\sigma,\mathcal{N}_{\rho_{SE}}} : X \mapsto \sigma^{1/2} \mathcal{N}^{\dagger}_{\rho_{SE}} (\mathcal{N}_{\rho_{SE}}(\sigma)^{-1/2} X \mathcal{N}_{\rho_{SE}}(\sigma)^{-1/2}) \sigma^{1/2}. \tag{5.12}$$

Since we here consider the trivial case where $\sigma' = \sigma$, the Petz map can always recover σ.

Unfortunately, in the case of interest here the existence of a Petz recovery map does not lead us to the sought-after unitary U_{SA}, since the Petz map doesn't necessarily take the form of a partial trace $\mathrm{Tr}_A(U_{SA}\,\rho'_{SA}\,U_{SA}^\dagger)$. Indeed, we can in fact argue that the Petz map *cannot* identically be the map $\mathrm{Tr}_A(U_{SA}\,\rho'_{SA}\,U_{SA}^\dagger)$. This latter recovery map cannot be CP over the set of all density operators if A, S, E are in an entangled state, which will generally be the case[3], whereas the Petz map is CP by construction. So while U_{SA} may exist, it cannot be found in this way.

In light of this difficulty, a different line of attack is to use the given unitary structure of the SCMERA as a starting point and see whether recovery can be engineered. This amounts to interpreting recovery as an instance of quantum error correction that protects against deletion of E. Think of the state σ_S that the ancillae are initialized in as an encoded message. At any given time step, the message is encoded into the combined SEA system by entangling it with EA. A part of the system, E, subsequently becomes inaccessible to us. We then wish to recover the encoded message by acting on the reduced SA system only with U_{SA}. If this is to be possible, then the allowed interactions U_{SE} are constrained. (This picture is reminiscent of quantum secret sharing.)

Since σ_S is the message that we want to recover and since we discard E, here $\mathcal{N}_{\rho_{SE}}$ is essentially a noisy channel, which we suppose takes on a particular Kraus form,

$$\mathcal{N}_{\rho_{SE}} : X \mapsto \sum_\mu N_\mu X N_\mu^\dagger, \tag{5.13}$$

for a given initial state ρ_{SE}[4]. In this context, in order for a recovery map R to exist, the Kraus operators N_μ must obey the following necessary and sufficient condition [152]. For the sake of generality, suppose that instead of wanting to recover a fixed state $|0\rangle_S$, the encoded message was chosen from a fixed subspace of S that has an orthonormal basis $\{|\phi_i\rangle_S\}$ (the specific case for SCMERA corresponds to there only being one basis vector, namely, $|0\rangle_S$). Then, the Kraus operators must obey the

[3]Even if one fixes a particular input at $t = 0$ to be a product state, entanglement will still be generated at a later time. This is because S, E generically become entangled after the isometry.

[4]Recall that any trace-preserving channel on a reduced system can be written using a (potentially input-dependent) set of Kraus operators $\{N_\mu\}$, where $\sum_\mu N_\mu^\dagger N_\mu = I$ [151].

Knill-Laflamme condition,

$$\langle \phi_i | N_\mu^\dagger N_\nu | \phi_j \rangle = C_{\mu\nu} \delta_{ij} \,, \tag{5.14}$$

where $C_{\mu\nu}$ is a Hermitian matrix. This condition places a constraint on what U_{SE} are allowed.

In the case of a single fixed state $|0\rangle_S$, the condition above is trivially satisfied, and so recovery is always possible. However, here as well it is not guaranteed whether there is a quantum error correcting code (QECC) on the whole SEA system that is consistent with SCMERA such that the ancillary state can always be recovered on the SA subsystem on the horizon. We do not know whether such a code exists, but it would have to satisfy certain requirements that we now explore.

In the case where the ancillary qudit is fixed to be a particular state, the code subspace is 1-dimensional. An implementation that allows one to decode the message may be possible to realize with the help of a $k = 0$ code[5]. (See [152] for a detailed review.) For a binary MERA, in which the interior, horizon, and ancillary Hilbert spaces are altogether comprised of 8 qudits, a satisfactory encoding would require a $[\![8, k, d]\!]$ code, where $k = 0$ if the ancillary states are always fixed to be $|0\rangle_S$. Because 2 qudits are effectively erased in discarding the interior (i.e., a known erasure location), the distance of the code must satisfy $d \geq t + 1$ with $t = 2$. As a zeroth order check, we see that this requirement is consistent with the quantum Singleton bound

$$n - k \geq 2(d - 1) \tag{5.15}$$

for $3 \leq d \leq 5$ with $k = 0$. Also note that, while we mainly consider the case where $k = 0$, larger code spaces with $k > 0$ (i.e., a situation where the ancillary state is chosen among several options at each step) are not ruled out. For example, a hypothetical tensor network that encodes $k = 2$ qudits worth of information could realize a QECC with $d = 3, 4$. We note that there exist binary codes that are compatible with our requirements on n, k, and d, for example, the $[\![8, 3, 3]\!]$ code (see

[5]The properties of a quantum error correcting code on qudits of dimension χ are often abbreviated by the notation $[\![n, k, d]\!]$ where n is the block size, k is the number of encoded qudits, and d the code distance. For $k = 0$, the χ^k-dimensional code subspace is precisely one dimensional.

section 7.12.3 in [153]), and presumably there also exist codes for qudit systems; however, we are unaware of their specific forms, and much less whether or not they are compatible with the tensor structure of SCMERA.

In summary, by interpreting SCMERA as a recovery operation or an error correcting code, we identify several necessary but generally insufficient criteria that the SCMERA circuit must meet. Note, however, that failure to meet these criteria cannot rule out strong complementarity, but it can rule out SCMERA as a model.

Finally, we elaborate a bit more on the unitarity of the proposed SCMERA circuit. The overall SCMERA tensor network can be understood as a circuit by including the ancillary degrees of freedom, S. In the case of perfect recovery of the ancillary state on the horizon, the ancillary state that was added in the interior can be discarded from the horizon at the end of the computation in each time step so that the total size of Hilbert space remains constant throughout. Alternatively, we can also understand the adding-and-discarding process as recycling the ancillary degrees of freedom at each step. It is clear in this sense that we have a unitary process on the same finite-dimensional Hilbert space. However, note that the unitary recovery mapping on the horizon need not recover the ancillary state perfectly. In fact, a universal (i.e., constant in time) unitary recovery map applied to every time step cannot in general achieve perfect recovery. In this case, recycling of the approximately recovered ancillary qudit will lead to information backflow into the static patch interior, which in turn leads to Poincaré recurrences. Discarding such ancillary qudits on the horizon avoids recurrences even when using a universal recovery map, but breaks unitarity. If we demand perfect recovery of the ancillary qudit, then the unitary evolution is necessarily time-dependent. It is, however, unclear if such time-dependence is only limited to swapping operations on the horizon.

5.5 Circuit Complexity and de Sitter Action

In AdS/CFT, the "complexity equals action" proposal [115] suggests that the complexity of a CFT thermofield double state as it evolves in time is proportional to the

Einstein-Hilbert (EH) action of a region of the bulk known as the Wheeler-De Witt patch. Explicitly, $\mathcal{C} = qS_{\rm EH}$, where the proportionality constant is calculated to be $q = 1/\pi\hbar$. Similarly, here we can show that complexity, calculated using the MERA circuit, scales in the same way as the corresponding spacetime action in de Sitter space.

For a given MERA-like circuit that is translationally and scale invariant, it is possible to estimate its complexity by choosing a reference state and gate set. It is natural to choose the reference state to be the initial state of dS/MERA, which we write as $|\Psi(t=0)\rangle = |\psi\rangle \otimes |\phi\rangle^{\otimes N}$. $|\Psi\rangle$ consists of the initial entangled component $|\psi\rangle$ which encodes the entanglement information needed to reconstruct the de Sitter spatial geometry at $t = 0$, and $|\phi\rangle^{\otimes N}$ denotes all the ancillary degrees of freedom that will later get entangled up to some time $t = T$. Here, because we only consider bounds on complexity, the estimate won't depend on the particular form of $|\psi\rangle$; we can take it to be an arbitrary state that lives on the initial few sites of the MERA at $t = 0$.

We obtain a straightforward estimate of complexity if we choose a reference gate set that corresponds to the exact disentanglers and isometries, $\{U, V\}$, that were used to build the MERA circuit. For a k-nary MERA, suppose that U, V are k-local and denote the total number of ancillae that get entangled up to time $t \leq T$ by

$$N(T) = \sum_{j=0}^{T} k^j. \tag{5.16}$$

It then follows that for any non-trivially entangled state $|\Psi(T)\rangle$, where none of the qudits in $|\Psi(T)\rangle$ can be written as a product state between the qudit and its complement[6], a lower bound on its complexity $\mathcal{C}(T)$ is proportional to $N(T)$. This is because, even using an optimal circuit that could potentially be more efficient than the MERA, it takes at least $N(T)/k$ k-local gates to even minimally entangle all of the product ancillae. The actual complexity to create the state with the correct entanglement structure at $t = T$ is therefore strictly lower-bounded. In addition, the MERA circuit itself that constructs the state $|\Psi(T)\rangle$ constitutes a trivial complexity upper bound. Hence, for generic scale and translationally invariant MERA

[6]For example, this is expected for a CFT vacuum state.

in arbitrary dimensions with k-local disentanglers and isometries, the complexity satisfies

$$C_0 N(T) \leq \mathcal{C}(T) \leq C_1 N(T), \tag{5.17}$$

where $C_1 > C_0$ are order-unity numbers that depend on the specific circuit construction. For the (1+1)-dimensional binary MERA shown, $C_0 = 4$ and $C_1 = 8$. Choosing a different reference gate set would give different coefficients C_0 and C_1, but the exponential dependence on T would remain unchanged.

An important distinction from the usual holographic complexity proposal [115] is the lack of a boundary theory, and hence a notion of bulk-boundary duality. Similarly, the proposal also differs from [154], where the complexity of the state on the de Sitter boundary is compared to the action or volume of a holographic asymptotically anti-de Sitter bulk. Because only the de Sitter bulk is present, we test a bulk complexity-action (volume) proposal by directly comparing the circuit complexity of MERA, which is conjectured to describe de Sitter spacetime, to the Einstein-Hilbert action (spacetime volume) of the same region in de Sitter.

The Einstein-Hilbert action of the portion of de Sitter spacetime covered by the global time interval $0 \leq t \leq T$ in D dimensions is given by

$$\begin{aligned} S_{\text{EH}} &= \frac{1}{16\pi G} \int_0^T dt \int d\Omega_{D-1} \sqrt{-g} R \\ &= \frac{R \ell_{\text{dS}}^D \mathcal{S}_{D-1}}{16\pi G} \int_0^T dt \, \cosh^{D-1} t \\ &= \frac{R \ell_{\text{dS}}^D \mathcal{S}_{D-1}}{16\pi G} \frac{1}{(D-1)2^{D-1}} e^{(D-1)T} \ + \ \text{subleading}, \end{aligned}$$

where $R = D(D-1)/\ell_{\text{dS}}^2 = 2D\Lambda/(D-2)$ is the Ricci curvature for de Sitter space with cosmological constant Λ and $\mathcal{S}_{D-1}$ is the volume of the $(D-1)$-sphere. We see that the scaling behavior is indeed consistent with the circuit complexity computed above, and the action satisfies the complexity bound for some appropriate choice of constant q. Note that each tensor in the MERA is mapped to a proper volume in de Sitter [105]. Therefore, comparison of other spacetime regions would yield a similar conclusion. It cannot differentiate the complexity = volume versus complexity =

action proposal, because the constant Ricci curvature in de Sitter space only changes q by a constant factor.

The proportionality constant between complexity and action depends on the choice of gate set, and differs from $q = 1/\pi\hbar$ in the original proposal. See [155, 156] for similar conclusions from more detailed studies in the context of quantum field theory. Interestingly, assuming the validity of the conjecture, the $(\ell_{\rm dS}/\ell_{\rm pl})^{D-2}$ scaling behavior in the action may suggest that the complexity of a correct circuit with sub-Hubble features should approximately scale as the horizon area (recall that R scales like $\ell_{\rm dS}^{-2}$). In the case of the MERA, this is encoded in the otherwise arbitrary choice of q, because the network structure is not sensitive to $\ell_{\rm dS}/\ell_{\rm pl}$.

5.6 Discussion

Discretizing de Sitter spacetime using the MERA seems to provide some interesting interpretations, in particular in terms of giving a natural information-theoretic reason for cosmic no-hair, constraining de Sitter complementarity, and giving the de Sitter action an information-theoretic interpretation. It would be interesting to ask what other consequences thinking of de Sitter spacetime in a tensor network/information-theoretic way could provide. For example, would a different tensor network discretization be more natural for answering other questions, or is the choice of tensor network discretization fixed by the spacetime metric one is attempting to duplicate? If so, are there other natural spacetimes (Lorentzian or Euclidean), for which different tensor networks might provide insights into open problems?

The MERA is naturally suited to describing de Sitter spacetime on super-Hubble scales, since structure within a horizon volume is not resolved. The state within a patch can nevertheless be encoded in the tensors inside the horizon, and perturbations of such a state in the de Sitter background can be initialized in the MERA input state. The cosmic no-hair result is then the fact that such perturbations flow to a fixed-point of the evolution superoperator within a patch.

Another limitation of this de Sitter-MERA correspondence is that it clearly breaks

the rotational symmetry of spacelike sections of de Sitter; a binary MERA that corresponds to (1+1)-dimensional de Sitter spacetime picks out four preferred causal patches, or equivalently, fixes the cardinal directions on the circle. It also breaks boost symmetry in that the MERA fixes a preferred global $t = 0$ slice. To this end, hyperinvariant tensor networks may be an interesting improvement on the MERA [107]. Hyperinvariant tensor networks were introduced to address, among other issues, a similar problem for AdS-MERA correspondences that the MERA picks out a preferred center point of the hyperbolic plane. In a hyperinvariant tensor network, any node in the tensor network can be taken to be the "center" of the hyperbolic plane, thus restoring a significant amount of symmetry. Since the radial direction in AdS corresponds to the renormalization direction of the MERA, which here corresponds to the timelike direction of de Sitter, a hyperinvariant tensor-network/de Sitter correspondence would likely no longer fix a preferred global $t = 0$ slice. Instead, the effective causal cone of any pair of adjacent nodes could be used to define a de Sitter static patch.

It would be interesting to push the present analysis beyond a strict de Sitter background. For example, it should be possible to adapt the tensor network to allow for bubble nucleation and eternal inflation. A classical variant of this was already considered in [157], and it would be useful to further investigate the evolution of quantum states using the kind of methods explored here.

Acknowledgements

We would like to thank Aleksander Kubica, Yasunori Nomura, Jason Pollack, and Grant Remmen for helpful discussions. This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of High Energy Physics, under Award Number DE-SC0011632, as well as by the Walter Burke Institute for Theoretical Physics at Caltech and the Foundational Questions Institute.

5.7 Stationary causal cones of the MERA

Given a k-nary MERA, in which the number of sites in each layer increases k-fold with every fine-graining step, what is the number of sites per layer of a stationary causal cone?

Example 5.7.1. Consider a binary MERA as in Fig. 5.1. Within the MERA, consider a set of sites at some layer and draw their causal cone in the coarse-graining direction. If the smallest simply-connected region that contains all of the initial sites is made up of L sites, then after $\sim \log_2 L$ steps in the coarse-graining direction, the causal cone will contain 2 or 3 sites [87]. Once the cone at some layer contains 2 or 3 sites, Fig. 5.9 illustrates how the width of the causal cone can evolve under further coarse-graining. Notice that if the cone contains 2 sites at some layer, then it is possible for the next layer to have either 2 or 3 sites, but if a given layer contains 3 sites, then all subsequent layers will contain 3 sites. Therefore, a stationary causal cone having the same width at every layer can only have 2 sites per layer or 3 sites per layer. In particular, only the stationary causal cone with 2 sites per layer is left/right-symmetric in a binary MERA.

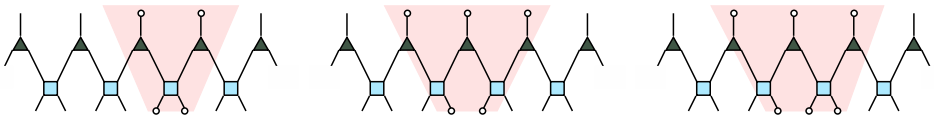

Figure 5.9: Ways in which a minimal-width causal cone can propagate between layers in a binary MERA. (a) $2 \to 2$, (b) $2 \to 3$, (c) $3 \to 3$.

Example 5.7.2. Consider a ternary MERA as in Fig. 5.4. Similarly, the causal cone of any given collection of sites will contain 2 or 3 sites after $\sim \log_3 L$ steps in the coarse-graining direction. If the cone contains 3 sites at some layer, then it is possible for the next layer to have either 2 or 3 sites, but if a given layer contains 2 sites, then all subsequent layers will contain 2 sites (Fig. 5.10). Therefore, a stationary causal cone having the same width at every layer can only have 2 sites per layer

or 3 sites per layer. Here, only the stationary causal cone with 3 sites per layer is left/right-symmetric in a ternary MERA.

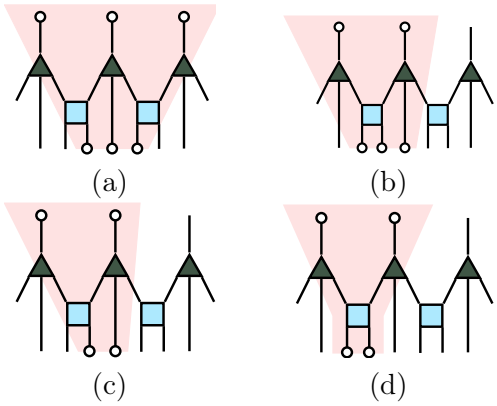

Figure 5.10: Ways in which a minimal-width causal cone can propagate between layers in a ternary MERA. (a) $3 \to 3$, (b) $3 \to 2$, (c) $2 \to 2$, first instance, (d) $2 \to 2$, second instance.

The case of a general k-nary MERA follows straightforwardly from the two examples above:

Proposition 5.7.3. *A stationary causal cone having the same width at every layer in a homogeneous k-nary MERA has 2 or 3 sites per layer.*

Proof: Given some homogeneous k-nary MERA with any arrangement of disentanglers and isometries, all of the legs in the tensor network can be blocked together to form composite legs so that the network takes the form of a binary or ternary MERA, as illustrated in Fig. 5.11, whence the proposition follows from the examples above. □

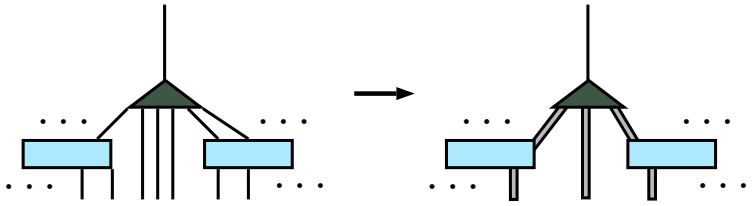

Figure 5.11: Legs in an arbitrary MERA can be blocked together. In this way, that the causal structure matches that of a binary or ternary MERA becomes apparent.

5.8 Higher-dimensional generalizations

Consider a d-dimensional MERA, where each layer is a hypercubic d-dimensional lattice. Here, the MERA is k-nary when each site in one layer gives rise to k^d sites in the next layer (see Fig. 5.12). The global MERA-de Sitter correspondence does not carry through in this case, simply because, on the de Sitter side, there is no way to latticize the d-sphere using a regular hypercubic lattice that is self-similar under fine-graining (although see [158] for a generalization to 2 dimensions).

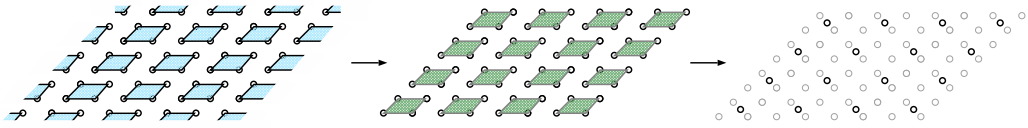

Figure 5.12: A 2D MERA. In a single coarse-graining step, blocks of 4 sites are acted on by a disentangler (blue), then blocks of 4 sites that are displaced from the last set of blocks are acted on by an isometry (green), reducing the number of sites by a factor of 4.

This is not to say that a generalization to higher dimension is impossible. One could consider a different tiling of global de Sitter that preserves uniformity and is self-similar under some refinement operation. For example, on a 2-sphere, regular or semi-regular tilings are possible using triangularizations, but these different tilings would necessarily require some sort of variation on the MERA tensor network. To the best of our knowledge, such generalizations are still unexplored.

On the other hand, one could still study the correspondence between de Sitter and a hypercubic MERA by restricting one's attention to only a single static patch. In this scenario, it is consistent to think of the MERA as defining a superoperator which maps the state on m^d sites of a given slice of a single static patch to the next slice. (Remember, the number of sites per horizon volume, i.e., per slice of the static patch, remains constant.) Therefore, the usual unmodified MERA may still be useful for understanding local aspects of de Sitter quantum gravity in higher dimensions.

Chapter 6

COSMIC EQUILIBRATION: A HOLOGRAPHIC NO-HAIR THEOREM FROM THE GENERALIZED SECOND LAW

In a wide class of cosmological models, a positive cosmological constant drives cosmological evolution toward an asymptotically de Sitter phase. Here we connect this behavior to the increase of entropy over time, based on the idea that de Sitter spacetime is a maximum-entropy state. We prove a cosmic no-hair theorem for Robertson-Walker and Bianchi I spacetimes that admit a Q-screen ("quantum" holographic screen) with certain entropic properties: If generalized entropy, in the sense of the cosmological version of the Generalized Second Law conjectured by Bousso and Engelhardt, increases up to a finite maximum value along the screen, then the spacetime is asymptotically de Sitter in the future. Moreover, the limiting value of generalized entropy coincides with the de Sitter horizon entropy. We do not use the Einstein field equations in our proof, nor do we assume the existence of a positive cosmological constant. As such, asymptotic relaxation to a de Sitter phase can, in a precise sense, be thought of as cosmological equilibration.

This chapter was published as Ref. [5], S. M. Carroll and A. Chatwin-Davies, "Cosmic equilibration: A holographic no-hair theorem from the generalized second law," Phys. Rev. D ***97*** *(2018) 046012,* `arXiv:1703.09241`.

6.1 Introduction

Like black holes, universes have no hair, at least if they have a positive cosmological constant Λ [119–128]. A cosmic no-hair theorem states that, if a cosmological spacetime obeys Einstein's equation with $\Lambda > 0$, then the spacetime asymptotically tends to an empty de Sitter state in the future.[1] A more precise statement is due to Wald, who proved the following theorem [119]:

[1] For a different definition of cosmic hair which more closely parallels black hole hair, see [159].

Theorem 6.1.1 (Wald)**.** *All Bianchi spacetimes (except for certain type IX spacetimes) that are initially expanding, that have a positive cosmological constant $\Lambda > 0$, and whose matter content besides Λ obeys the strong and dominant energy conditions, tend to a de Sitter state in the future.*

Bianchi spacetimes are cosmologies that are homogeneous but in general anisotropic [160, 161]. For example, the metric of the 1+3-dimensional Bianchi I spacetime in comoving Cartesian coordinates is given by

$$ds^2 = -dt^2 + a_1^2(t)\,dx^2 + a_2^2(t)\,dy^2 + a_3^2(t)\,dz^2. \tag{6.1}$$

It is essentially a Robertson-Walker (RW) spacetime in which the scale factor can be different in different directions in space. In this case, when the necessary conditions are satisfied, Wald's theorem implies that each $a_i(t)$ tends to the same de Sitter scale factor, $\exp(\sqrt{\Lambda/3}\,t)$ for a cosmological constant $\Lambda > 0$, as t tends to infinity.

The intuition behind why one would expect a cosmic no-hair theorem to hold is that as space expands, the energy density of ordinary matter decreases while the density of vacuum energy remains constant. As such, the cosmological constant eventually dominates regardless of the initial matter content and geometry, and a universe in which a positive cosmological constant is the only source of stress-energy is de Sitter. For Bianchi I spacetimes, one can make this intuition explicit by writing down a Friedmann equation for the average scale factor, $\bar{a}(t) \equiv [a_1(t)a_2(t)a_3(t)]^{1/3}$, which gives [162, Ch. 8.6]

$$\left(\frac{\dot{\bar{a}}(t)}{\bar{a}(t)}\right)^2 \propto \left(\rho_\Lambda + \rho_{\text{matter}} + \rho_{\text{an}}\right). \tag{6.2}$$

On the right-hand side, ρ_Λ and ρ_{matter} denote the energy densities due to the cosmological constant and matter respectively, while ρ_{an} is an effective energy density due to anisotropy, similar to how one can think of spatial curvature as an effective source of stress-energy. Crucially, ρ_{an} scales at most like $\bar{a}^{-2}$, and so as the universe expands, only the constant contribution due to ρ_Λ persists. The exception to Wald's theorem is the case of a Bianchi IX spacetime (which has positive spatial

curvature) whose initial matter energy density is so high that the spacetime recollapses before the cosmological constant can dominate [119]. Intuitively, we expect not only anisotropies, but also perturbative inhomogeneities to decay away at late times, though this is harder to prove rigorously [120, 127, 163, 164]. For arbitrary inhomogeneous and anisotropic cosmologies, one can always find regions that expand at least as fast as de Sitter, thus realizing a type of local no-hair theorem [165]. Beyond classical general relativity, various generalizations of Wald's theorem attempt to demonstrate analogous no-hair theorems for the quantum states of fields on a curved spacetime background [129–131].

As the universe expands and the cosmological constant increases in prominence with respect to other energy sources, something else is also going on: entropy is increasing. According to the Second Law of Thermodynamics, the entropy of any closed system (such as the universe) will increase or stay constant, at least until it reaches a maximum value. It is interesting to ask whether there is a connection between these two results, the cosmic no-hair theorem and the Second Law. Can the expansion of the universe toward a quiescent de Sitter phase be interpreted as thermodynamic equilibration to a maximum-entropy state? It is well established that de Sitter has many of the properties of an equilibrium maximum-entropy state, including a locally thermal density matrix with a constant temperature [116, 166], and the relationship between entropy and de Sitter space has been examined from a variety of perspectives [112, 137–141, 167–169].

In this paper we try to make one aspect of these ideas rigorous, showing that a cosmic no-hair theorem can be derived even without direct reference to Einstein's equation, simply by invoking an appropriate formulation of the Second Law. This strategy of deducing properties of spacetime from the behavior of entropy is reminiscent of the thermodynamic and entropic gravity programs [93,170–173], as well as of the gravity-entanglement connection [20, 24, 65–67, 76, 86, 174]. Though we do not attempt to derive a complete set of gravitational field equations from entropic considerations, it is interesting that a specific spacetime can be singled out purely from the requirement

that entropy increases to a maximum finite value.

To derive our theorem, we require a precise formulation of the Second Law that is applicable in curved spacetime, and that includes the entropy of spacetime itself. A step in this direction is Bekenstein's Generalized Second Law (GSL) [13]. Recall that the entropy of a black hole with area A is given by $S_{\rm BH} = A/4G$. The GSL is the conjecture that *generalized entropy*, $S_{\rm gen}$, which is defined as the sum of the entropy of all black holes in a system as well as the ordinary thermodynamic entropy, increases or remains constant over time. Unfortunately this form of the GSL does not immediately help us in spacetimes without any black holes. Recently, Bousso and Engelhardt proposed a cosmological version of the GSL [29], building on previous work on holography [175], apparent horizons [176–181], and holographic screens [182, 183]. They define a version of generalized entropy on a hypersurface they call a "Q-screen." A Q-screen is a quantum version of a holographic screen, which in turn is a modification of an apparent horizon. Given a Cauchy hypersurface Σ and a codimension-2 spatial surface with no boundary $\sigma \subset \Sigma$ that divides Σ into an interior region and an exterior region, the generalized entropy is the sum of the area entropy of σ, i.e., its area in Planck units, and the entropy of matter in the exterior region:

$$S_{\rm gen}[\sigma, \Sigma] = \frac{A[\sigma]}{4G} + S_{\rm out}[\sigma, \Sigma]. \tag{6.3}$$

Bousso and Engelhardt's version of the GSL is the statement that generalized entropy increases strictly monotonically with respect to the flow through a specific preferred foliation of a Q-screen:

$$\frac{dS_{\rm gen}}{dr} > 0\,, \tag{6.4}$$

where r parameterizes the foliation. Although it is unproven in general, this version of the GSL is well motivated and known to hold in specific circumstances (the discussion of which we defer to the next section).

In this work, we use the GSL to establish a cosmic no-hair theorem purely on thermodynamic grounds. In an exact de Sitter geometry, the de Sitter horizon is a holographic screen[2], and every finite horizon-sized patch is associated with a fixed

[2]Pure de Sitter spacetime does not, however, satisfy the generic conditions outlined in [183].

entropy that is proportional to the area of the horizon in Planck units [184]. We therefore conjecture that evolution toward such a state is equivalent to thermodynamic equilibration of a system with a finite number of degrees of freedom, and therefore a finite maximum entropy. Specifically, assuming the GSL, we show that if a Bianchi I spacetime admits a Q-screen along which generalized entropy monotonically increases up to a finite maximum, then the anisotropy necessarily decays and the scale factor approaches de Sitter behavior asymptotically in the future. At no point do we use the Einstein field equations, nor do we assume the presence of a positive cosmological constant. The GSL and that entropy tends to a finite maximum along the Q-screen take the logical place of these two respective ingredients.

The proof essentially consists of first showing that an approach to a finite maximum entropy heavily constrains the possible asymptotic structure of a Q-screen. Second, we show that the spacetime must necessarily be asymptotically de Sitter (and in particular, isotropic as well) in order to admit a Q-screen with the aforementioned asymptotic structure.

The structure of the rest of this paper is as follows. We review Q-screens and the GSL in Sec. 6.2. In Sec. 6.3, we first prove a cosmic no-hair theorem for the simpler case of RW spacetimes using the GSL. Then, in Sec. 6.4, we move on to the proof for Bianchi I spacetimes, first in 1+2 dimensions to illustrate our methods, and then in 1+3 dimensions, which also illustrates how to generalize to arbitrary dimensions. We discuss aspects of the theorems and their proofs as well as some implications in Sec. 6.5.

6.2 The generalized second law for cosmology

We begin by briefly reviewing Bousso and Engelhardt's conjectured Generalized Second Law (GSL). The GSL can be thought of as a quasilocal version of Bekenstein's entropy law for black holes [13], but which also applies to cosmological settings. Moreover, the GSL is a natural semiclassical extension of Bousso and Engelhardt's area theorem for holographic screens in the same way that Bekenstein's entropy law

extends Hawking's area theorem to evaporating black holes.

An early cornerstone of classical black hole thermodynamics [185,186] was Hawking's area theorem: in all spacetimes which satisfy the null curvature condition, the total area of all black hole event horizons can only increase, i.e., $dA/dt \geq 0$ [187]. Of course, the area theorem fails for evaporating black holes, the technical evasion being that they do not satisfy the null curvature condition. Bekenstein pointed out, however, that if one instead interprets the area of the black hole event horizon as horizon entropy and includes the entropy of the Hawking radiation outside the black hole, S_{out}, in the total entropy budget, then the generalized entropy, $S_{\text{gen}} = A/4G + S_{\text{out}}$, increases monotonically or stays constant, $dS_{\text{gen}}/dt \geq 0$ [13].

From the perspective of trying to understand the thermodynamics of spacetime, both Hawking's and Bekenstein's results suffer from two inconveniences. First, they are fundamentally nonlocal, since identifying event horizons requires that one know the full structure of a Lorentzian spacetime. Second, these results only apply to black holes; it would be desirable to understand thermodynamic aspects of spacetime in other geometries as well. These considerations motivate holographic screens [182, 183], a subset of which obey a classical area theorem, as well as their semiclassical extensions called Q-screens [29], a subset of which are conjectured to obey an entropy theorem. Importantly, both holographic screens and Q-screens are quasilocally defined and are known to be generic features of cosmologies in addition to black hole spacetimes.

Let us first review holographic screens. Following the convention of Bousso and Engelhardt, here and throughout we will refer to a spacelike codimension-2 hypersurface simply as a "surface."

Let σ be a compact connected surface. At every point on σ, there are two distinct future-directed null directions (or equivalently, two distinct past-directed null directions) that are orthogonal to σ: inward- and outward-directed. The surface σ is said to be *marginal* if the expansion of the null congruence corresponding to one of these directions, say k^μ, is zero everywhere on σ. Consequently, σ is a slice of the null sheet

generated by k^μ that locally has extremal area. This last point is particularly clear if one observes that the expansion, $\theta = \nabla_\mu k^\mu$, at a point $y \in \sigma$, can be equivalently defined as the rate of change per unit area of the area of the slice, $A[\sigma]$, when a small patch of proper area $\mathcal{A}$ is deformed along the null ray generated by k^μ at y with an affine parameter λ:

$$\theta(y) = \lim_{\mathcal{A}\to 0} \frac{1}{\mathcal{A}} \frac{dA[\sigma]}{d\lambda}\bigg|_y \tag{6.5}$$

This definition is illustrated in Fig. 6.1 below.

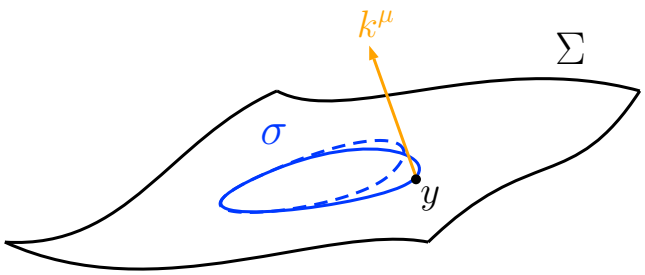

Figure 6.1: Given a Cauchy hypersurface Σ, the surface $\sigma \subset \Sigma$ (drawn with a solid line) splits Σ into an interior and exterior. Deformations of σ (drawn with a dotted line) are defined by dragging σ along the null ray generated by k^μ at any point $y \in \sigma$. More precisely, a deformation is defined by transporting a small area element $\mathcal{A} \subset \sigma$ at y in the k^μ direction.

A *holographic screen* is a smooth codimension-1 hypersurface that can be foliated by marginal surfaces, which are then called its *leaves*. Note that while the leaves σ are spacelike, in general a holographic screen need not have a definite character. A marginal surface σ is said to be *marginally trapped* if the expansion of the congruence in the other null direction is negative everywhere on σ, and a *future holographic screen* is a holographic screen whose leaves are marginally trapped; marginally anti-trapped surfaces and past holographic screens are defined analogously. Then, assuming the null curvature condition as well as a handful of mild generic conditions, Bousso and Engelhardt proved that future and past holographic screens obey the area theorem paraphrased below [182, 183]:

Theorem 6.2.1 (Bousso & Engelhardt)**.** *Let H be a regular holographic screen. The area of its leaves changes strictly monotonically under the flow through the foliation of H.*

Q-screens are related to holographic screens, but with expansion replaced by what is dubbed the "quantum expansion." Let σ again denote a compact connected surface. The quantum expansion at a point $y \in \sigma$ in the orthogonal null direction k^μ is defined as the rate of change per unit proper area of the generalized entropy (6.3), i.e., the sum of both area and matter entropy, with respect to affine deformations along the null ray generated by k^μ:

$$\Theta_k[\sigma; y] = \lim_{\mathcal{A}\to 0} \frac{4G}{\mathcal{A}} \frac{dS_{\text{gen}}}{d\lambda}\bigg|_y \tag{6.6}$$

Then similar to before, a *quantum marginal surface* is a surface σ such that the quantum expansion in one orthogonal null direction vanishes everywhere on σ. Just as a marginal surface locally extremizes area along a lightsheet, a quantum marginal surface locally extremizes the generalized entropy along the lightsheet generated by k^μ.

The adjective "quantum" can be confusing in this context. In this work it denotes a shift from classical general relativity, where one proves theorems about the area of surfaces, to quantum field theory on a semiclassical background, where analogous theorems refer to a generalized entropy that adds the entropy of matter degrees of freedom to such an area. That matter entropy may be be calculated as the quantum (von Neumann) entropy of a density operator, but in the right circumstances (which we will in fact be dealing with below) it is equally appropriate to treat it as a classical thermodynamic quantity. So here "quantum" should always be interpreted as "adding an entropy term to the area of some surface," whether or not quantum mechanics is directly involved.

The remaining constructions have similarly parallel definitions. A *Q-screen* is a smooth codimension-1 hypersurface that can be foliated by quantum marginal surfaces. A quantum marginal surface σ is *marginally quantum trapped* if the quantum expansion in the other null direction is negative everywhere on σ, and a *future Q-screen* is a Q-screen whose leaves are marginally quantum trapped. Analogous definitions apply for anti-trapped surfaces and past Q-screens. A Q-screen may be timelike, null, spacelike, or some combination thereof in different regions. Future and

past Q-screens that also obey certain generic conditions analogous to those for holographic screens are the objects that are conjectured to obey a Generalized Second Law [29]:

Conjecture 6.2.2 (Generalized Second Law)**.** *Let $\mathcal{Q}$ be a regular future (resp. past) Q-screen. The generalized entropy of its leaves increases strictly monotonically under the past and outward (resp. future and inward) flow along $\mathcal{Q}$.*

Note that while the GSL remains unproven in general, it is known to hold in several examples, and it can in fact be shown to hold if one assumes the Quantum Focusing Conjecture [34].

So far we have not said much about the precise definition of generalized entropy, so let us discuss how it is defined in more careful terms. Our context here is quantum field theory in curved spacetime, rather than a full-blown theory of quantum gravity. Given some spacetime, suppose that it comes equipped with a foliation by Cauchy hypersurfaces, and suppose that the spacetime's matter content is described by a density matrix $\rho(\Sigma)$ on each Cauchy hypersurface Σ. Let σ be a compact connected surface that divides a Cauchy hypersurface Σ into two regions: the interior and exterior of σ. The generalized entropy computed with respect to σ and Σ is then the sum of the area of σ in Planck units and S_{out}, the von Neumann entropy of the reduced state of ρ restricted to the exterior of σ, cf. Eq. (6.3). The reduced state of ρ outside σ, which we denote ρ_{out}, is obtained by tracing out degrees of freedom on Σ in the interior of σ,

$$\rho_{\text{out}} \equiv \text{tr}_{\text{int}\,\sigma}[\rho(\Sigma)]\,, \tag{6.7}$$

and the Von Neumann entropy of ρ_{out} is

$$S_{\text{out}}[\sigma, \Sigma] = -\text{tr}\left[\rho_{\text{out}} \ln \rho_{\text{out}}\right]. \tag{6.8}$$

For a general field-theoretic state, the von Neumann entropy $S_{\text{out}}[\sigma, \Sigma]$ is a formally divergent quantity. Consequently, there is some subtlety surrounding how it should be regulated, whether through an explicit ultraviolet cutoff or via subtracting a

divergent vacuum contribution [29, 95]. Since we will exclusively be concerned with cosmology, we will work in a regime where the matter content of the spacetime has a conserved "thermodynamic," or coarse-grained entropy s per unit comoving volume. (Entropy per comoving volume is approximately conserved in cosmologies that do not have too much particle production [162, Ch 3.4].) The von Neumann entropy of a quantum mechanical system coincides with the thermodynamic Gibbs entropy in the classical limit where the state ρ_{out} has no coherence, i.e., is diagonal in the energy eigenbasis of Gibbs microstates.

We will suppose that we can take the matter contribution to the generalized entropy, which is formally given by the von Neumann entropy $S_{\text{out}}[\sigma, \Sigma]$, to be given by a coarse-grained entropy $S_{\text{CG}}[\sigma, \Sigma]$ in the interior of σ:

$$S_{\text{out}}[\sigma, \Sigma] \;\rightarrow\; S_{\text{CG}}[\sigma, \Sigma] = s \cdot \text{vol}_c[\sigma, \Sigma]. \tag{6.9}$$

Here, $\text{vol}_c[\sigma, \Sigma]$ denotes the comoving (coordinate) volume of $\text{int}\,\sigma$ on Σ. (This approach is also taken in the examples of [29].) This expression is appropriate for cosmology, where observers find themselves on the inside of Q-screens and cosmological horizons when present, as opposed to observers who remain outside of a black hole and who are unable to access the interior of the black hole's horizon. Moreover, in the field-theoretic case where $\rho(\Sigma)$ is a pure state, then it follows that $S_{\text{in}} = S_{\text{out}}$, where S_{in} is the Von Neumann entropy of $\rho_{\text{in}} \equiv \text{tr}_{\text{ext}\,\sigma}[\rho(\Sigma)]$.

The fact that each leaf of a Q-screen extremizes the generalized entropy on an orthogonal lightsheet leads to a useful method for constructing Q-screens [175]. Given some spacetime with a foliation by Cauchy surfaces, suppose that one is also supplied with a foliation of the spacetime by null sheets with compact spatial cross-sections. Let each null sheet be labeled by a parameter r, and on each null sheet, let $\sigma(r)$ be the spatial section with extremal generalized entropy, when it exists. (Not every spacetime contains Q-screens, such as Minkowski space. But in Big Bang cosmologies, we expect both the area of, and entropy inside, a light cone to decrease in the very far past, so the generalized entropy will have an extremum somewhere.) It follows that each $\sigma(r)$ is a quantum marginal surface, and so if the quantum expansion

has a definite sign in the other orthogonal null direction on each $\sigma(r)$, the union of these surfaces, $\mathcal{Q} = \bigcup_r \sigma(r)$, is by construction a Q-screen.

One way to generate a null foliation of a spacetime is to consider the past light cones of some timelike trajectory. Q-screens constructed from this type of foliation will be particularly useful for our purposes. This construction is illustrated through a worked example in Section 6.6.

6.3 A cosmic no-hair theorem for RW spacetimes

We can used the notions reviewed above to show that spacetimes that expand and approach a constant maximum entropy along Q-screens will asymptote to de Sitter space. The basic idea of our proof is made clear by the simple example of a metric that is already homogeneous and isotropic, so that all we are showing is that the scale factor approaches e^{Ht} for some fixed constant H. The anisotropic case, considered in the next section, is considerably more complex, but the ideas are the same.

Let $\mathcal{M}$ be a Robertson-Walker (homogeneous and isotropic) spacetime with the line element

$$ds^2 = -dt^2 + a^2(t)\left(d\chi^2 + \chi^2 d\Omega_{d-1}^2\right), \tag{6.10}$$

where $t \in (t_i, \infty)$. Our aim is to show that if $\mathcal{M}$ admits a past Q-screen along which the generalized entropy monotonically increases up to a finite maximum value, then this alone, together with a handful of generic conditions on $\mathcal{M}$, implies that $\mathcal{M}$ is asymptotically de Sitter, or in other words, that

$$\lim_{t\to\infty} a(t) = e^{Ht} \tag{6.11}$$

for some constant H. In particular, we will neither make use of the Einstein field equations nor assume that there is a positive cosmological constant.

Begin by foliating $\mathcal{M}$ with past-directed light cones whose tips lie at the spatial origin $\chi = 0$, and suppose that $\mathcal{M}$ admits a past Q-screen, $\mathcal{Q}$, constructed with respect to this foliation. In other words, suppose that each light cone has a spatial slice with extremal generalized entropy so that $\mathcal{Q}$ is the union of all of these extremal

slices. Past light cones will generically have a maximal entropy slice in cosmologies which, for example, begin with a big bang where $a(t_i) = 0$. An example is portrayed in Fig. 6.2, which shows a holographic screen and a Q-screen in a cosmological spacetime with a past null singularity and a future de Sitter evolution; this example is explained in more detail in Section 6.6. The intuition here is that while the past-directed null geodesics that make up a light cone may initially diverge, eventually they must meet again in the past when the scale factor vanishes and space becomes singular. Ultimately, however, we need only assume that the Q-screen exists, and we only remark on its possible origins for illustration.

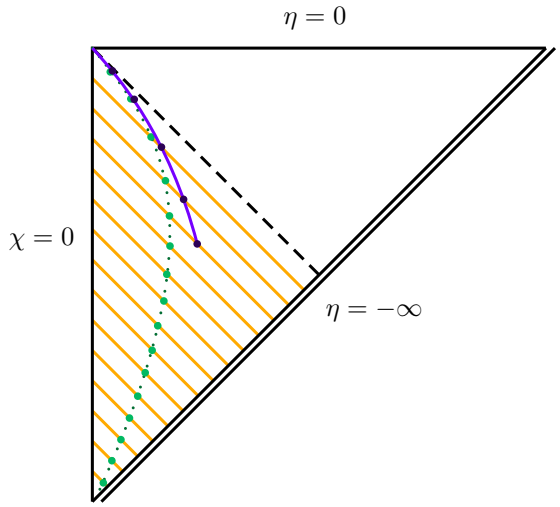

Figure 6.2: Holographic screen and Q-screen illustrated on the Penrose diagram for a homogeneous and isotropic spacetime with a positive cosmological constant. Null sheets that make up the foliation by past-directed light cones are shown in yellow, and the cosmological horizon is the dashed black line. The dotted green line and large green dots are the holographic screen and its leaves, respectively. The solid purple line and large purple dots are the Q-screen and its leaves.

Because RW spacetimes are spherically symmetric, the extremal-entropy light cone slices will be spheres, i.e., constant-t slices. If the quantum expansion vanishes in the lightlike direction along the light cone and is positive in the other lightlike direction at a single point on some test sphere, then it maintains these properties at every point on that sphere due to symmetry. This sphere is by construction a marginally quantum anti-trapped surface, or equivalently has extremal generalized entropy on the light cone. We therefore take the Cauchy surfaces Σ with respect to which

generalized entropy is defined to be the constant-t surfaces in $\mathcal{M}$, since constant-t slices of light cones are spheres.

We will also make a handful of generic assumptions about $\mathcal{M}$ and $\mathcal{Q}$ without which a cosmic no-hair theorem is not guaranteed. Indeed, Wald's theorem does not hold in completely general cosmologies either; it requires that the spacetime is initially expanding and that its matter content satisfies the strong and dominant energy conditions. Here, we will assume that space continues to expand for all cosmic time.[3] We want to avoid cosmologies that crunch or that otherwise clearly do not admit a no-hair theorem. We will also suppose that $\mathcal{Q}$ satisfies the generic conditions outlined in [29].

With these considerations in mind, the theorem that we wish to prove is the following:

Theorem 6.3.1. *Let $\mathcal{M}$ be a RW spacetime with the line element* (6.10) *and whose matter content has constant thermodynamic entropy s per comoving volume. Suppose that $\mathcal{M}$ admits a past Q-screen, $\mathcal{Q}$, constructed with respect to a foliation of $\mathcal{M}$ with past-directed light cones that are centered on the origin, $\chi = 0$, and suppose that the Generalized Second Law holds on $\mathcal{Q}$. Suppose that $\mathcal{M}$ and $\mathcal{Q}$ together satisfy the following assumptions:*

(*a*) $a(t) \to \infty$ *as* $t \to \infty$,

(*b*) $S_{\rm gen} \to S_{\rm max} < \infty$ *along* $\mathcal{Q}$.

Then, $\mathcal{M}$ is asymptotically de Sitter and the scale factor $a(t)$ approaches e^{Ht}, where H is a constant.

Proof: For convenience we work in $d = 3$ spatial dimensions, but the generalization to arbitrary dimensions is straightforward. As discussed above, the leaves of $\mathcal{Q}$ are spheres. Letting the leaves be labeled by some parameter r, the generalized entropy

[3] In principle, the expansion need not be monotonic, but we will find that monotonicity is implied when $\mathcal{M}$ admits a Q-screen such as $\mathcal{Q}$.

is then given by

$$S_{\text{gen}}[\sigma(r), \Sigma(r)] \equiv S_{\text{gen}}(r) = \frac{\pi}{G}\chi^2(r)a^2(t(r)) + \tfrac{4}{3}\pi\chi^3(r)s\,. \tag{6.12}$$

The hypersurface $\Sigma(r)$ is the constant-$t(r)$ surface in which the leaf $\sigma(r)$ is embedded, and $\chi(r)$ denotes the radius of the leaf.

First, we need to establish that $\mathcal{Q}$ extends out to future timelike infinity. In principle, $\mathcal{Q}$ could become spacelike and consequently not extend beyond some time t (or in other words, $t(r)$ could have some finite maximum value), but it turns out that this does not happen.

Recall the property of Q-screens that generalized entropy is extremized on each leaf with respect to lightlike deformations. Here we may write

$$k^\mu \partial_\mu S_{\text{gen}} = 0\,, \tag{6.13}$$

where $k^\mu = (a(t), -1, 0, 0)$ is the lightlike vector that is tangent to the light cone and with respect to which S_{gen} is extremal. (Any point x^μ belongs to a unique sphere on a past-directed light cone and may therefore be associated with a particular value of S_{gen}. This lets us define the partial derivative in Eq. (6.13) above.) The deformation corresponds to dragging the leaf $\sigma(r)$ up and down the light cone, and by construction $S_{\text{gen}}(r)$ is extremal on the leaf $\sigma(r)$. Note that in more general settings we should consider deformations with respect to null *geodesics*, since the null generators of the light cone could have different normalizations at different points on $\sigma(r)$. Or, in other words, the geometry of the leaf $\sigma(r)$ could change as it is dragged by some fixed affine amount along the light cone. Here, however, the spherical symmetry of RW ensures that the null generators on $\sigma(r)$ all have the same normalization, so that k^μ as defined above is proportional to the null generators everywhere on $\sigma(r)$.

Writing out the partial derivatives, (6.13) becomes

$$\begin{aligned} 0 &= (a\,\partial_t - \partial_\chi)\left(\frac{\pi}{G}\chi^2 a^2 + \tfrac{4}{3}\pi\chi^3 s\right) \\ &= \frac{2\pi}{G}\chi^2 a^2 \dot{a} - \frac{2\pi}{G}\chi a^2 - 4\pi\chi^2 s. \end{aligned} \tag{6.14}$$

(One must be careful to distinguish between the coordinate t and the value $t(r)$ which labels leaves in the Q-screen.) If $\chi \neq 0$, then it follows that

$$\frac{1}{\chi} = \dot{a}(t) - \frac{2Gs}{a^2(t)}\,. \tag{6.15}$$

Eq. (6.15) lays out the criterion for when there is a leaf in a constant-t slice; when the right side is finite and positive, then there must be a leaf in that slice.

Observe that the right side of Eq. (6.15) does not diverge for any finite $t > t_i$ since $a(t)$ is defined for all $t \in [t_i, \infty)$ and only diverges in the infinite t limit by assumption. Furthermore, if the right side is nonzero and positive for some time t_{time} (and consequently there is a leaf $\sigma(r)$ in the $t(r) = t_{\text{time}}$ slice), then the right side cannot approach zero, since this would cause the radius of subsequent leaves to grow infinitely large, which contradicts the assumption that S_{gen} remains finite. Therefore, if $\mathcal{Q}$ has a leaf at *some* time, then Eq. (6.15) shows that $\mathcal{Q}$ must have leaves in *all* future slices. $\mathcal{Q}$ is therefore timelike and extends out to future timelike infinity.[4] Furthermore, that the right side of Eq. (6.15) cannot vanish immediately implies that $\dot{a} > 2Gs/a^2 > 0$ for $t > t_{\text{time}}$, so that the expansion must be monotonic.

Because $\mathcal{Q}$ is timelike, we can label each leaf by the constant-t_1 surface in which it lies, i.e., let the parameter r be a time t_1 (subscripted as such to distinguish it from the coordinate t). Referring to Eq. (6.12), since $a(t)$ grows without bound by assumption, it must be that $\chi(t_1)$ decreases at least as fast as $a^{-1}(t_1)$ in order for the area term in S_{gen} to remain finite (as it must, since by hypothesis $S_{\text{gen}} \leq S_{\max}$). The matter entropy term therefore becomes irrelevant in the asymptotic future, and so that $S_{\text{gen}} \to S_{\max}$, it must be that

$$\chi(t_1) \to \sqrt{\frac{GS_{\max}}{\pi}}\frac{1}{a(t_1)} \tag{6.16}$$

as $t_1 \to \infty$.

[4] Alternatively, we could instead replace Assumption (a) with the assumption that $\mathcal{Q}$ is timelike and extending out to future timelike infinity and argue that $a \to \infty$. The arguments given here show that these two points are logically equivalent.

Next, rearrange Eq. (6.15) to solve for $\dot{a}$. Using the asymptotic form for $\chi(t_1)$ in Eq. (6.16), to leading order in a we find that

$$\dot{a} \to \sqrt{\frac{\pi}{GS_{\max}}}a \quad + \quad (\text{subleading})\,. \tag{6.17}$$

Therefore, it follows that $a(t) \to e^{Ht}$ as $t \to \infty$, where $H = (\pi/GS_{\max})^{1/2}$, demonstrating that the metric approaches the de Sitter form, as desired. The entropy $S_{\max} = \pi/GH^2$ coincides with the usual de Sitter horizon entropy. □

We close this section by briefly remarking that the result above extends straightforwardly to open and closed RW spacetimes as well.

Corollary 6.3.2. *More generally, the result of Theorem 6.3.1 applies to a RW spacetime $\mathcal{M}$ of any spatial curvature, i.e., with the line element*

$$ds^2 = -dt^2 + a^2(t)\left(d\chi^2 + f^2(\chi)d\Omega_{d-1}^2\right), \tag{6.18}$$

where

$$f(\chi) = \begin{cases} \sin\chi & \chi \in [0,\pi] \quad (\text{closed}) \\ \chi & \chi \in [0,\infty) \quad (\text{flat}) \\ \sinh\chi & \chi \in [0,\infty) \quad (\text{open}) \end{cases}. \tag{6.19}$$

Proof: The overall proof technique is the same as in the proof of Theorem 6.3.1. Working in 1+3 dimensions, in the more general case, the generalized entropy of the leaves of $\mathcal{Q}$ is given by

$$S_{\text{gen}}[\sigma(r), \Sigma(r)] \equiv S_{\text{gen}}(r) = \frac{\pi}{G}f^2(\chi(r))a^2(t(r)) + v(\chi(r))s\,. \tag{6.20}$$

When $\mathcal{M}$ is closed, the comoving volume $v(\chi)$ is given by $v(\chi) = 2\pi(\chi - \sin\chi\cos\chi)$, and when $\mathcal{M}$ is open, $v(\chi)$ is given by $v(\chi) = 2\pi(\sinh\chi\cosh\chi - \chi)$. Consequently, the condition $k^\mu\partial_\mu S_{\text{gen}} = 0$, which determines when there is a leaf in the constant-t hypersurface, gives

$$\frac{1}{f^2(\chi)} = \left(\dot{a}(t) - \frac{2Gs}{a^2(t)}\right)^2 + k\,, \tag{6.21}$$

where $k = +1, 0$, or -1 if $\mathcal{M}$ is respectively closed, flat, or open. Here as well, if there is a leaf at some time t_{time} so that the right-hand side of Eq. (6.21) is nonzero, then there are leaves in all subsequent constant-t slices, since the finiteness of S_{gen} demands that the right-hand side cannot approach zero. Therefore, $\mathcal{Q}$ extends out to future timelike infinity.

For the general case, the condition in Eq. (6.16) that $S_{\text{gen}} \to S_{\text{max}}$ reads[5]

$$f(\chi(t_1)) \to \sqrt{\frac{GS_{\text{max}}}{\pi}} \frac{1}{a(t_1)} \,. \tag{6.22}$$

Upon substituting Eq. (6.22) into Eq. (6.21) (and taking the positive root, since $\mathcal{M}$ is expanding), we recover Eq. (6.17), and so the rest of the proof follows as before.

□

6.4 A cosmic no-hair theorem for Bianchi I spacetimes

In a RW spacetime, we demonstrated that the existence of a Q-screen along which entropy monotonically increases to a finite maximum implies that the scale factor tends to the de Sitter scale factor far in the future. Now we will go one step further and show that in the case where the cosmology is allowed to be anisotropic, similar assumptions imply that any initial anisotropies decay at late times as well. Specifically, we will prove a cosmic no-hair theorem for Bianchi I spacetimes. The calculations in the proof for Bianchi I spacetimes are more involved than the RW case, so we will begin with a proof in 1+2 dimensions, where the anisotropy only has one functional degree of freedom. We will then generalize to 1+3 dimensions, which also makes apparent how to generalize to arbitrary dimensions.

6.4.1 1+2 dimensions

Let $\mathcal{M}$ be a Bianchi I spacetime in 1+2 dimensions with the line element

$$ds^2 = -dt^2 + a_1^2(t)\, dx^2 + a_2^2(t)\, dy^2, \tag{6.23}$$

[5] A minor technical point worth noting is that the condition in Eq. (6.22) is not identically equivalent to the condition $\chi(t_1) \to \sqrt{GS_{\text{max}}/\pi}/a(t_1)$ when $\mathcal{M}$ is closed. In this case, $\chi(t_1) \to \pi - \sqrt{GS_{\text{max}}/\pi}/a(t_1)$ is also admissible.

where $t \in (t_i, \infty)$. Once again foliate $\mathcal{M}$ with past-directed light cones whose tips lie at $x = y = 0$, and suppose that $\mathcal{M}$ admits a past Q-screen $\mathcal{Q}$, constructed with respect to this foliation, together with an accompanying foliation by Cauchy hypersurfaces. Our aim is to show that if generalized entropy tends to a finite maximum along $\mathcal{Q}$, then the GSL implies that $a_1(t), a_2(t) \to e^{Ht}$ as $t \to \infty$ for some constant H.

Here as well we will assume that space expands for all time, with $a_1(t), a_2(t) \to \infty$ as $t \to \infty$. We will also further assume that $\mathcal{Q}$ is timelike and extends out to future timelike infinity past some time t_{time}. We suspect that it might be possible to show that this latter property follows from the assumption that $a_1(t)$ and $a_2(t)$ grow without bound, as in the case of a RW spacetime, but we do not know of a straightforward way to show this.

We will assume that generalized entropy is *globally* maximized on each light cone by the corresponding screen leaf (as opposed to only assuming local extremality). In other words, we will assume that there are no other slices of each light cone whose generalized entropy is larger than that of the screen leaf. This property of leaves is certainly true when the Quantum Focusing Conjecture (QFC) holds [34]. Moreover, the GSL is provably true when the QFC holds.

The QFC is the conjecture that the quantum expansion of a null congruence is nonincreasing along the congruence. In symbols, for a null congruence generated by k^μ with an affine parameter λ on a given null ray, the QFC reads

$$\frac{d\Theta_k}{d\lambda} \leq 0\,. \tag{6.24}$$

The QFC is the semiclassical analogue of classical focusing, $d\theta/d\lambda \leq 0$, which holds when the null curvature condition holds. In particular, Eq. (6.24) makes it clear that if a light cone slice locally maximizes generalized entropy with respect to deformations on the light cone, then it is also the unique global maximum. A leaf σ that locally maximizes generalized entropy obeys $\Theta_k[\sigma, y] = 0$ for all $y \in \sigma$. Therefore, if

Θ_k is nonincreasing on the light cone[6], there are no deformations of σ that lead to a larger generalized entropy, and so σ defines a globally maximal generalized entropy. It is interesting to explore ways in which this assumption about global maximality of generalized entropy can be relaxed, which we shall do after the proof of the no-hair theorem.

Next, we introduce *conformal light cone coordinates* [188, 189], which are more convenient coordinates to work in when dealing with anisotropy. First, observe that we may rewrite the line element (6.23) as

$$ds^2 = -dt^2 + a^2(t)\left[e^{2b(t)}dx^2 + e^{-2b(t)}dy^2\right] \tag{6.25}$$

with $a_1(t) = a(t)\,e^{b(t)}$ and $a_2(t) = a(t)\,e^{-b(t)}$ [161]. In this parameterization, the "volumetric scale factor" $a(t)$ describes the overall expansion of space while $b(t)$ characterizes the anisotropy. Next, make the coordinate transformation to conformal time defined by $dt = \pm a(\eta)\,d\eta$ so that the line element (6.25) reads

$$ds^2 = a^2(\eta)\left[-d\eta^2 + e^{2b(\eta)}dx^2 + e^{-2b(\eta)}dy^2\right]. \tag{6.26}$$

Choose the sign of η so that $\eta(t)$ is a monotonically increasing function of t, and denote the limiting value of $\eta(t)$ as $t \to \infty$ by η_∞. Conformal light cone coordinates are then defined by the following coordinate transformation:

$$x(\eta, \eta_o, \theta) = \cos\theta \int_\eta^{\eta_o} \frac{e^{-2b(\zeta)}}{\sqrt{\cos^2\theta\ e^{-2b(\zeta)} + \sin^2\theta\ e^{2b(\zeta)}}}\ d\zeta \tag{6.27}$$

$$y(\eta, \eta_o, \theta) = \sin\theta \int_\eta^{\eta_o} \frac{e^{2b(\zeta)}}{\sqrt{\cos^2\theta\ e^{-2b(\zeta)} + \sin^2\theta\ e^{2b(\zeta)}}}\ d\zeta. \tag{6.28}$$

The point with coordinates (η, η_o, θ) is reached by firing a past-directed null geodesic from the spatial origin $x = y = 0$ at an angle $\theta \in [0, 2\pi)$ counterclockwise relative to the x-axis at conformal time η_o and following the light ray in the past down to the conformal time η (Fig. 6.3). Note that while η is a timelike coordinate, η_o acts as a radial coordinate at each η.

[6]The pathological case of $d\Theta_k/d\lambda = 0$ on a subset of the congruence with nonzero measure is ruled out by appropriate genericity conditions.

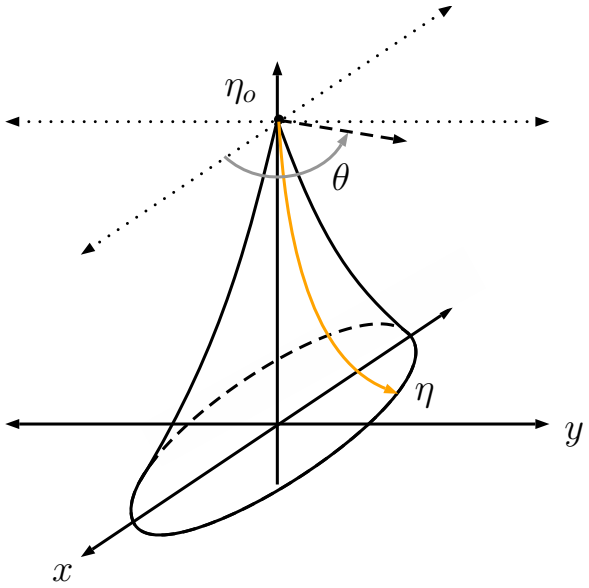

Figure 6.3: Conformal light cone coordinates. At the conformal time η_o, fire a past-directed null geodesic (shown in yellow) from the origin at an initial angle θ relative to the positive x-axis and follow it until the conformal time η.

The surfaces of constant η_o are precisely the past-directed light cones with respect to which $\mathcal{Q}$ is constructed. We can therefore label the leaves σ of $\mathcal{Q}$ by the values of η_o corresponding to the light cones on which they lie (Fig. 6.4):

$$\mathcal{Q} = \bigcup_{\eta_o} \sigma(\eta_o). \tag{6.29}$$

Similarly, label the Cauchy hypersurfaces with respect to which each leaf is defined by $\Sigma(\eta_o)$. In various instances, it will be useful to use another coordinate,

$$\chi = \eta_o - \eta, \tag{6.30}$$

which may be thought of as a comoving radius in a sense that will be made precise later. We will also sometimes work in the coordinates (η, χ, θ) or (χ, η_o, θ) in addition to the conformal light cone coordinates (η, η_o, θ).

The no-hair theorem that we will prove is as follows:

Theorem 6.4.1. *Let $\mathcal{M}$ be a Bianchi I spacetime with the line element* (6.23) *and whose matter content has constant thermodynamic entropy s per comoving volume. Suppose that $\mathcal{M}$ admits a past Q-screen $\mathcal{Q}$, with globally maximal entropy leaves*

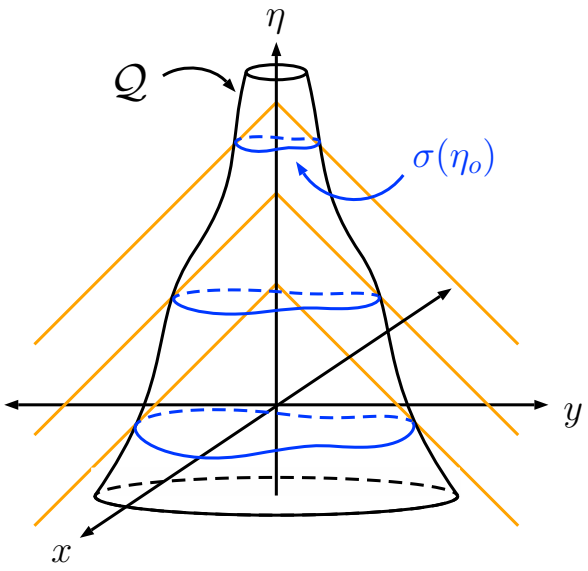

Figure 6.4: A Q-screen $\mathcal{Q}$ (the solid black hypersurface) constructed with respected to a foliation by past-directed light cones (sketched in yellow). Each leaf $\sigma(\eta_o)$ (shown in blue) is labelled by the value of η_o where the tip of its parent light cone sits.

constructed with respect to a foliation of $\mathcal{M}$ with past-directed light cones that are centered on the origin, $x = y = 0$. Suppose that the Generalized Second Law holds on $\mathcal{Q}$ and that $\mathcal{M}$ and $\mathcal{Q}$ together satisfy the following assumptions:

(*i*) $a_1(t), a_2(t) \to \infty$ *as* $t \to \infty$,

(*ii*) $\mathcal{Q}$ *is timelike past some* t_{time} *and extends out to future timelike infinity*

(*iii*) $\dot{a}_1(t), \dot{a}_2(t) > 0$ *after some* t_{mono},

(*iv*) $S_{\text{gen}} \to S_{\text{max}} < \infty$ *along* $\mathcal{Q}$.

Then, $\mathcal{M}$ is asymptotically de Sitter and the scale factors $a_1(t)$ and $a_2(t)$ approach $C_1 e^{Ht}$ and $C_2 e^{Ht}$, respectively, where H, C_1, and C_2 are constants.

Notes: To obtain a manifestly isotropic metric, rescale the coordinates x and y by C_1 and C_2, i.e., set $X = C_1 x$ and $Y = C_2 y$. Then, the line element (6.23) asymptotically reads $ds^2 \to -dt^2 + e^{2Ht}(dX^2 + dY^2)$. Also note that we have introduced an additional assumption compared to the RW case: Assumption (*iii*),

that $a_1(t)$ and $a_2(t)$ grow monotonically past some time t_{mono}. Finally, also note that in terms of $a(\eta)$ and $b(\eta)$, Assumption (i) becomes:

(i') $a(\eta) \to \infty$ as $\eta \to \eta_\infty$ and $a(\eta)e^{\pm b(\eta)} \to \infty$.

In terms of $a(\eta)$ and $b(\eta)$, the theorem is established by showing that $a(\eta) \to -1/H\eta$ and $b(\eta) \to B$ as $\eta \to 0^-$ (and also that $\eta_\infty = 0$) for some constant B.

Proof: The proof can be broken down into three parts. First, we show that, asymptotically, $\mathcal{Q}$ squeezes into the comoving coordinate origin. Second, we use this asymptotic squeezing behaviour to demonstrate that the volumetric scale factor $a(\eta)$ tends to the de Sitter scale factor. Finally, we show that the asymptotic behaviour of $a(\eta)$ and Assumption (iii) together imply that anisotropy decays.

Showing that $\mathcal{Q}$ squeezes into the coordinate origin $\chi = 0$ as $\eta \to \eta_\infty$.

Consider the leaves of $\mathcal{Q}$ and work in $\tilde{x}^\mu = (\eta, \eta_o, \theta)$ coordinates. On the light cone whose tip is at η_o, each leaf $\sigma(\eta_o)$ is a closed path parameterized by

$$\tilde{x}^\mu(u; \eta_o) = (\eta(u; \eta_o), \eta_o, u) \qquad u \in [0, 2\pi). \tag{6.31}$$

Our first task is to show that $\chi(u; \eta_o) \equiv \eta_o - \eta(u; \eta_o)$ tends to zero for all values of u as $\eta \to \eta_\infty$. We will do so through a proof by contradiction.

Suppose to the contrary that $\mathcal{Q}$ never squeezes into the comoving coordinate origin. That is, suppose that there exists $M > 0$ such that, given any $\eta_o > \eta_{\text{time}}$, one can find values $\tilde{\eta}_o > \eta_o$ and $\tilde{u}$ such that $\chi(\tilde{u}; \tilde{\eta}_o) \geq M$. Let $\tilde{\eta} \equiv \eta(\tilde{u}; \tilde{\eta}_o)$ and consider the constant $\eta = \tilde{\eta}$ slice of the light cone whose tip is at $\tilde{\eta}_o$ (Fig. 6.5). Denote this (co-dimension 2) surface by $\varsigma(\tilde{\eta}; \tilde{\eta}_o)$, and denote the (co-dimension 1) hypersurface of constant-$\tilde{\eta}$ by $X(\tilde{\eta})$. Since the generalized entropy of the leaf $\sigma(\tilde{\eta}_o)$ is globally maximal on this light cone by assumption, it must follow that

$$S_{\text{gen}}[\sigma(\tilde{\eta}_o), \Sigma(\tilde{\eta}_o)] \geq S_{\text{gen}}[\varsigma(\tilde{\eta}; \tilde{\eta}_o), X(\tilde{\eta})] \geq \frac{A[\varsigma(\tilde{\eta}; \tilde{\eta}_o)]}{4G}\,, \tag{6.32}$$

where the last inequality follows because $S_{\rm gen}$ is always greater than or equal to just the area term. Our basic strategy will be to show that $A[\varsigma(\tilde{\eta};\tilde{\eta}_o)]$ diverges as $\tilde{\eta}_o \to \eta_\infty$, which contradicts the assumption (iv) that $S_{\rm gen}$ must remain finite on $\mathcal{Q}$.

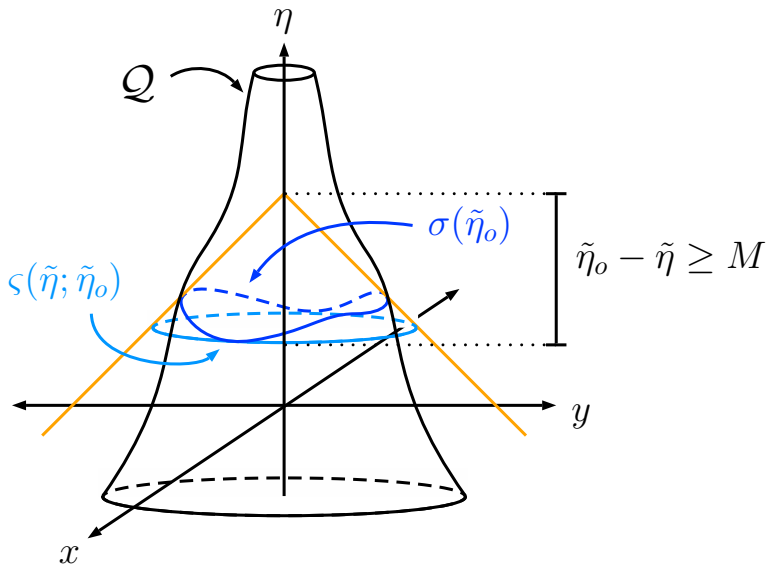

Figure 6.5: The leaf $\sigma(\tilde{\eta}_o)$ and the constant-$\tilde{\eta}$ slice, $\varsigma(\tilde{\eta};\tilde{\eta}_o)$, of its parent light cone.

To do this, let us compute the proper area $A[\varsigma(\tilde{\eta};\tilde{\eta}_o)]$. In three dimensions, the induced metric on a surface of constant η and η_o has only a single component, given by

$$\gamma = \frac{\partial x^\mu}{\partial\theta}\frac{\partial x^\nu}{\partial\theta} g_{\mu\nu} = a^2(\eta)\left[e^{2b(\eta)}\left(\frac{\partial x}{\partial\theta}\right)^2 + e^{-2b(\eta)}\left(\frac{\partial y}{\partial\theta}\right)^2\right] \equiv a^2(\eta)\,\tilde{\gamma}\,, \qquad (6.33)$$

where the coordinate partial derivatives read[7]

$$\frac{\partial x}{\partial\theta} = \int_\eta^{\eta_o} \frac{-\sin\theta}{\left(\cos^2\theta\; e^{-2b(s)} + \sin^2\theta\; e^{2b(s)}\right)^{3/2}}\, ds$$
$$\frac{\partial y}{\partial\theta} = \int_\eta^{\eta_o} \frac{\cos\theta}{\left(\cos^2\theta\; e^{-2b(s)} + \sin^2\theta\; e^{2b(s)}\right)^{3/2}}\, ds.$$

It follows that the area of this surface is

$$A(\eta,\eta_o) = \int_0^{2\pi}\sqrt{\gamma}\; d\theta = a(\eta)\int_0^{2\pi}\sqrt{\tilde{\gamma}}\; d\theta\,. \qquad (6.34)$$

[7] A Maple worksheet which implements the calculations in this article is available through the online repository [190].

It is fairly straightforward to place a lower bound on this area:

$$\begin{aligned} A(\eta,\eta_o) &\geq a(\eta)e^{b(\eta)}\int_0^{2\pi}\left|\frac{\partial x}{\partial\theta}\right|\,d\theta \\ &= a(\eta)e^{b(\eta)}\int_\eta^{\eta_o} ds\int_0^{2\pi} d\theta\frac{|\sin\theta|}{\left(\cos^2\theta\ e^{-2b(s)}+\sin^2\theta\ e^{2b(s)}\right)^{3/2}} \\ &= 4\,a(\eta)e^{b(\eta)}\int_\eta^{\eta_o} ds\ e^{-b(s)}. \end{aligned}$$

One arrives at a similar expression using $\partial y/\partial\theta$. Note that in the middle line above, we were able to bring the absolute value into the integrand of $\partial x/\partial\theta$ because it has a definite sign for any given θ. Then, if $e^{-b(s)}$ is minimized at $s=\eta_m\in[\eta,\eta_o]$, it follows that

$$A(\eta,\eta_o)\geq 4\,a(\eta)e^{b(\eta)}e^{-b(\eta_m)}(\eta_o-\eta)\geq 4\,a(\eta)(\eta_o-\eta). \tag{6.35}$$

Applied to our surface $\varsigma(\tilde\eta;\tilde\eta_o)$, for which $\tilde\eta_o-\tilde\eta\geq M$, the bound reads

$$A[\varsigma(\tilde\eta;\tilde\eta_o)]\equiv A(\tilde\eta,\tilde\eta_o)\geq 4Ma(\tilde\eta)\,, \tag{6.36}$$

which diverges as $\tilde\eta_o$ and $\tilde\eta$ are chosen arbitrarily large. We therefore have the contradiction that we sought, and so the leaves of the Q-screen must squeeze into the comoving coordinate origin in the asymptotic future.

Showing that $a(\eta)$ is asymptotically de Sitter.

Now we turn our attention to calculating $S_{\rm gen}[\sigma(\eta_o),\Sigma(\eta_o)]$ itself, and using its asymptotic properties as $\eta_o\to\eta_\infty$ to demonstrate that $a(\eta)\to -1/H\eta$ for a constant H with $\eta_\infty=0$. First, we we will argue that the matter entropy term, which we assume can be calculated using the coarse-grained entropy $S_{\rm CG}[\sigma(\eta_o),\Sigma(\eta_o)]$, vanishes asymptotically in the future. To this end, let us prove the following useful lemma about constant-η slices of light cones when $\chi=\eta_o-\eta$ is infinitesimally small:

Lemma 6.4.2. *Let $\varsigma(\eta;\eta+\chi)$ be the constant-η slice of the past-directed light cone whose tip is at $\eta_o=\eta+\chi$. The generalized entropy defined by this slice is given by*

$$S_{\rm gen}[\varsigma(\eta;\eta+\chi),X(\eta)]=\frac{A(\eta,\eta+\chi)}{4G}+c_g(\eta,\chi)\chi^2 s, \tag{6.37}$$

where $A(\eta, \eta+\chi)$ is given by

$$A(\eta, \eta+\chi) = a(\eta) \cdot \left[2\pi\chi + O(\chi^3)\right], \tag{6.38}$$

and $c_g(\eta, \chi)$ is some $O(1)$ geometric factor due to anisotropy that does not depend on $a(\eta)$.

Proof: First we justify the parameterization of the coarse-grained entropy $S_{\rm CG} = c_g(\eta,\chi)\chi^2 s$. In the coordinates of the metric (6.26), $S_{\rm CG}$ is given by

$$S_{\rm CG}[\varsigma(\eta;\eta+\chi), X(\eta)] = s \cdot {\rm vol}_c[\varsigma(\eta;\eta+\chi), X(\eta)] = s \iint_{{\rm int}\,\varsigma} dx\, dy\,, \tag{6.39}$$

where ${\rm int}\,\varsigma(\eta;\eta+\chi)$ denotes the region on $X(\eta)$ inside $\varsigma(\eta;\eta+\chi)$. In terms of the coordinates (η,χ,θ), $S_{\rm CG}$ is

$$S_{\rm CG}[\varsigma(\eta;\eta+\chi), X(\eta)] \equiv S_{\rm CG}(\eta,\chi) = s\int_0^{\chi}\int_0^{2\pi} \left|\frac{\partial(x,y)}{\partial(\chi',\theta)}\right| \, d\theta\, d\chi'\,. \tag{6.40}$$

Formally, the Jacobian can be calculated from the coordinate transformation (6.27)-(6.28) above. Expanding in powers of χ, one finds that

$$S_{\rm CG}(\eta,\chi) = s\cdot\left(\pi\chi^2 + \frac{\pi}{8}b'(\eta)^2\chi^4\right) + O(\chi^5)\,. \tag{6.41}$$

Therefore, we can simply define the function $c_g(\eta,\chi)$ to be the function

$$c_g(\eta,\chi) \equiv \frac{S_{\rm CG}(\eta,\chi)}{\chi^2 s} = \pi + \frac{\pi}{8}b'(\eta)^2\chi^2 + O(\chi^3)\,. \tag{6.42}$$

The function $c_g(\eta,\chi)$ is $O(\chi^0)$ by construction, and from the coordinate transformation (6.27)-(6.28), in which $a(\eta)$ never appears, we see that c_g cannot depend on $a(\eta)$, as claimed.

The expansion of $A(\eta,\eta+\chi)$ for small χ follows from expanding $\sqrt{\widetilde{\gamma}}$ in Eq. (6.34) in powers of χ and then integrating. Note that Eq. (6.38) demonstrates the sense in which χ is a comoving radius (at least for small values of χ). ■

We can use the result of Lemma 6.4.2 to show that $S_{\rm CG}[\sigma(\eta_o), \Sigma(\eta_o)]$ vanishes asymptotically in the future. Given a leaf $\sigma(\eta_o)$, let $\eta_{\rm min}$ be the minimum value attained

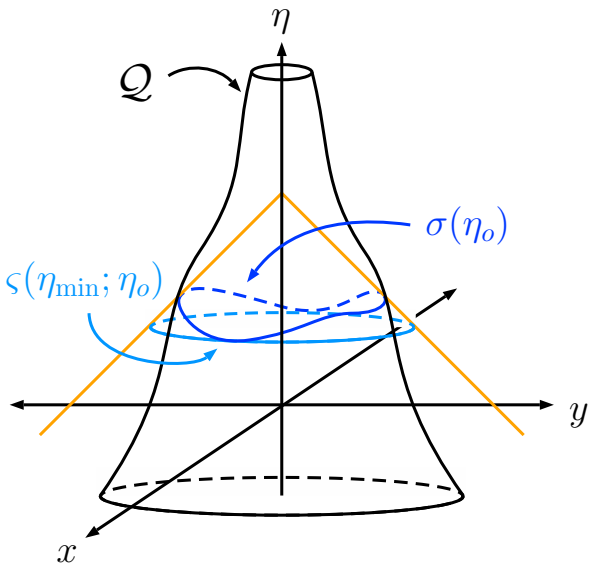

Figure 6.6: Given a leaf $\sigma(\eta_o)$, the constant-$\eta = \eta_{\rm min}$ slice of its parent light cone is the surface $\varsigma(\eta_{\rm min};\eta_o)$.

by $\eta(u;\eta_o)$:

$$\eta_{\rm min} = \min_u \{\eta(u;\eta_o)\}\,. \tag{6.43}$$

Consider the constant-$\eta_{\rm min}$ slice of the light cone whose tip is at η_o, which we label by $\varsigma(\eta_{\rm min};\eta_o)$ (Fig. 6.6). The comoving volume of $\sigma(\eta_o)$ is contained within the comoving volume of $\varsigma(\eta_{\rm min};\eta_o)$, which, according to Lemma 6.4.2, vanishes in the asymptotic future limit. Therefore, the comoving volume of $\sigma(\eta_o)$ vanishes as well, so $S_{\rm CG}[\sigma(\eta_o),\Sigma(\eta_o)]$ vanishes asymptotically in the future.

Next we investigate the asymptotic behaviour of $A[\sigma(\eta_o)]$. For this part of the proof, we will work in the coordinates (χ,η_o,θ). In these coordinates, the leaf $\sigma(\eta_o)$ is parameterized by some path $\tilde{x}^\mu(u) = (\chi(u;\eta_o),\eta_o,u)$ with η_o held constant and $0 \le u < 2\pi$. In the future when $S_{\rm CG}$ becomes negligible, this path is the maximal area (also known as length in 1+2 dimensions) path on the light cone whose tip is at η_o, and so $A[\sigma(\eta_o)]$ satisfies

$$\frac{\delta A[\sigma(\eta_o)]}{\delta\chi(u;\eta_o)} = 0\,. \tag{6.44}$$

In principle, one can therefore solve the Euler-Lagrange problem above to obtain the path $\chi(u;\eta_o)$ and hence also the maximal area $A[\sigma(\eta_o)]$.

The tangent to the path is $t^\mu = d\tilde{x}^\mu/du = (\dot{\chi}(u;\eta_o), 0, 1)$ (where a dot denotes a derivative with respect to the parameter u). Therefore, the area of $\sigma(\eta_o)$ is given by

$$A[\sigma(\eta_o)] = \int_0^{2\pi} \sqrt{\tilde{g}_{\mu\nu} t^\mu t^\nu}\, du = \int_0^{2\pi} \sqrt{\tilde{g}_{00}\dot{\chi}^2 + 2\tilde{g}_{02}\dot{\chi} + \tilde{g}_{22}}\, du, \tag{6.45}$$

where $\tilde{g}_{\mu\nu}$ is the metric of Eq. (6.26) but rewritten in (χ, η_o, θ) coordinates. One finds that $\tilde{g}_{00} = 0$ exactly, but $\tilde{g}_{02}$ and $\tilde{g}_{22}$ do not admit any such simplifications. Because of this, solving the full Euler-Lagrange problem to actually obtain the path $\chi(u;\eta_o)$ is intractable in general.

Nevertheless, we can exploit the fact that $\mathcal{Q}$ squeezes into the coordinate origin and perform a small-χ expansion of $A[\sigma(\eta_o)]$. First, pull out a factor of $a(\eta_o - \chi)$ from the square root:

$$A[\sigma(\eta_o)] = \int_0^{2\pi} a(\eta_o - \chi)\sqrt{2f_{02}\dot{\chi} + f_{22}}\, du\,. \tag{6.46}$$

In so doing we have defined $\tilde{g}_{\mu\nu} = [a(\eta_o - \chi)]^2 f_{\mu\nu}$. Then, expand the square root in χ. The result is

$$A[\sigma(\eta_o)] = \int_0^{2\pi} a(\eta_o - \chi)\left[\frac{\chi}{R(u;\eta_o)} + \frac{1}{2}b'(\eta_o)\frac{Q(u;\eta_o)}{R^2(u;\eta_o)}\chi^2 + O(\chi^3)\right] du, \tag{6.47}$$

where

$$\begin{aligned} R(u;\eta_o) &= e^{-2b(\eta_o)}\cos^2 u + e^{2b(\eta_o)}\sin^2 u \\ Q(u;\eta_o) &= e^{-2b(\eta_o)}\cos^2 u - e^{2b(\eta_o)}\sin^2 u. \end{aligned} \tag{6.48}$$

Pulling out the scale factor is necessary to avoid pathologies that arise because both χ and η_o become small in the same limit (see Section 6.6 for illustration).

Only keeping the first order term, the variation $\delta A/\delta\chi = 0$ gives

$$0 = -a'(\eta_o - \chi)\frac{\chi}{R(u;\eta_o)} + a(\eta_o - \chi)\frac{1}{R(u;\eta_o)}\,, \tag{6.49}$$

so asymptotically, the maximizing path $\chi(u;\eta_o) = \chi(\eta_o)$ is given implicitly by the solution of

$$\chi = \frac{a(\eta_o - \chi)}{a'(\eta_o - \chi)}. \tag{6.50}$$

To first order, $A[\sigma(\eta_o)]$ is given by

$$A[\sigma(\eta_o)] = 2\pi \frac{a^2(\eta_o - \chi)}{a'(\eta_o - \chi)}. \tag{6.51}$$

But the requirement that $S_{\rm gen} \to S_{\rm max}$ means that $A[\sigma(\eta_o)]/4G$ must tend to the constant value $S_{\rm max}$, or in other words,

$$\lim_{\substack{\eta_o \to \eta_\infty \\ \chi \to 0}} \frac{a^2(\eta_o - \chi)}{a'(\eta_o - \chi)} = \frac{2GS_{\rm max}}{\pi} \equiv \frac{1}{H}\,. \tag{6.52}$$

Therefore, $a(\eta)$ asymptotically approaches de Sitter, $a(\eta) \to -1/H\eta$ as $\eta \to 0^-$, with $H = \pi/2GS_{\rm max}$.

Since $\chi(\eta_o)$ is a function of η_o, a technical detail to address is to check that the higher-order coefficients in the expansion (6.47), which themselves depend on η_o through $b(\eta_o)$ and its derivatives, do not cause the higher-order terms to be larger than the term that is first-order in χ. This we can achieve by bounding the remainder, $r_1(\chi;\eta_o) \equiv \sqrt{2f_{02}\dot{\chi} + f_{22}} - \chi/R$.

Let $F = \sqrt{2f_{02}\dot{\chi} + f_{22}}$. We may write its second derivative with respect to χ as

$$\frac{\partial^2 F}{\partial \chi^2} = b'(\eta_o - \chi)\frac{Q(u;\eta_o - \chi)}{R^2(u;\eta_o - \chi)} + \varepsilon(\chi;\eta_o), \tag{6.53}$$

where the term $\varepsilon(\chi;\eta_o) \to 0$ as $\chi \to 0$ for any η_o. As such, choose χ and η_o both small enough such that $|\varepsilon(\chi;\eta_o)| < |b'(\eta_o - \chi)|/R(u;\eta_o - \chi)$ for all u.[8] With this choice, and since $|Q/R| \leq 1$, we have that

$$\left|\frac{\partial^2 F}{\partial \chi^2}\right| < \frac{2|b'(\eta_o - \chi)|}{R(u;\eta_o - \chi)}\,. \tag{6.54}$$

Next we invoke the monotonicity Assumption (*iii*). Let $\eta_\star = \max\{\eta_{\rm mono}, \eta_{\rm time}\}$. In terms of $a(\eta)$ and $b(\eta)$, Assumption (*iii*) reads $(a(\eta)e^{+b(\eta)})' > 0$ and $(a(\eta)e^{-b(\eta)})' > 0$, or $|b'(\eta)| < a'(\eta)/a(\eta)$. Therefore, upon additionally requiring $0 > \eta_o - \chi > \eta_\star$ (i.e. possibly making χ and η_o smaller), it follows that

$$\left|\frac{\partial^2 F}{\partial \chi^2}\right| < \frac{2}{R(u;\eta_o - \chi)}\frac{a'(\eta_o - \chi)}{a(\eta_o - \chi)} \;\longrightarrow\; \frac{2}{R(u;\eta_o - \chi)\chi(\eta_o)}\,. \tag{6.55}$$

[8] The only instance in which this is not possible is if $|b'(\eta_o - \chi)|/R(u;\eta_o - \chi)$ vanishes faster than $|\varepsilon(\chi;\eta_o)|$. But, in this case, the remainder $|r_1(\chi;\eta_o)|$ can be bounded arbitrarily tightly, since $|\partial^2 F/\partial\chi^2|$ can be made arbitrarily small.

So, by Taylor's remainder theorem we have that $|r_1(\chi;\eta_o)| < R(u;\eta_o-\chi_1)^{-1}(\chi^2/\chi(\eta_o))$ on any interval $\chi \in [\chi(\eta_o),\chi_2]$, where $\chi_1 \in [\chi(\eta_o),\chi_2]$ minimizes $R(u;\eta_o-\chi)$. Or, at the edge of the interval,

$$|r_1(\chi(\eta_o);\eta_o)| < \frac{\chi(\eta_o)}{R(u;\eta_o-\chi_1)}\,. \tag{6.56}$$

Since $\int_0^{2\pi} R(u;\eta_o)^{-1}\,du = 2\pi$ irrespective of the value of η_o, it follows that remainder in the expansion is strictly smaller than the first-order term, so we were safe in restricting our attention to the first-order solution.

Showing that the anisotropy decays.

The decay of anisotropy is directly implied by Assumption (*iii*) once we have established that the volumetric scale factor $a(\eta)$ is asymptotically de Sitter. In the far future limit, Assumption (*iii*) recast as $(a(\eta)e^{\pm b(\eta)})' > 0$ gives

$$|b'(\eta)| < \frac{a'(\eta)}{a(\eta)} \overset{\eta\to 0^-}{\longrightarrow} Ha(\eta) = \frac{1}{-\eta}\,. \tag{6.57}$$

Therefore, to capture the asymptotic scaling of $b'(\eta)$, we can write

$$b'(\eta) = \frac{f(\eta)}{(-\eta)^{1-p}}, \tag{6.58}$$

where $p > 0$ and where $|f(\eta)| \leq F$ for some bounded constant F when $\eta > \eta_\star$. In other words, $b'(\eta)$ cannot grow faster than $1/\eta$ as $\eta \to 0^-$, so that $(-\eta)^{1-p}b'(\eta)$ is some bounded function. To establish that the anisotropy decays, and thus complete the proof of the theorem, we need only establish that $b(\eta)$ goes to a fixed limit at late times:

Lemma 6.4.3. *If $b'(\eta)$ satisfies Eq. (6.58) on $(\eta_\star, 0)$, then $\lim_{\eta\to 0^-} b(\eta)$ exists.*

Proof: We show that the limit exists by showing that $b(\eta)$ is a Cauchy function. Let $\epsilon > 0$. We must find $\delta > 0$ such that $0 < -\eta_1 < \delta$ and $0 < -\eta_2 < \delta$ implies that

$|b(\eta_2) - b(\eta_1)| < \epsilon$. Without loss of generality, suppose that $\eta_\star < \eta_1 < \eta_2$. Then:

$$\begin{aligned} |b(\eta_2) - b(\eta_1)| &= \left| \int_{\eta_1}^{\eta_2} \frac{f(u)}{(-u)^{1-p}} \, du \right| \\ &\leq \int_{\eta_1}^{\eta_2} \frac{|f(u)|}{(-u)^{1-p}} \, du \\ &\leq F \int_{\eta_1}^{\eta_2} \frac{1}{(-u)^{1-p}} \, du \\ &\leq \frac{F}{p} (-\eta_1)^p. \end{aligned} \tag{6.59}$$

Therefore, let $\delta = (\epsilon F/p)^{1/p}$. Then $|b(\eta_2) - b(\eta_1)| < \epsilon$ as required. ■

□

It is interesting to briefly consider how one may relax the assumption that each light cone has a globally maximal generalized entropy section.[9] If we do not assume that each light cone has a maximum generalized entropy surface, then the proof above pauses at Eq. (6.32). In this case, it is no longer true that $S_{\text{gen}}[\sigma(\tilde{\eta}_o), \Sigma(\tilde{\eta}_o)]$ must be greater than $S_{\text{gen}}[\varsigma(\tilde{\eta}; \tilde{\eta}_o), X(\tilde{\eta})]$; the generalized entropy of the leaf $\sigma(\tilde{\eta}_o)$ could just be a local maximum, and the entropy of the constant-$\tilde{\eta}$ slice $\varsigma(\tilde{\eta}; \tilde{\eta}_o)$ could be larger. We must therefore make a slightly different argument. It turns out that a weaker but sufficient assumption is to only assume that each light cone has a unique maximum *area* surface.

As before, let us still suppose that $\mathcal{Q}$ never squeezes into the comoving coordinate origin and find a contradiction. We again suppose that there exists $M > 0$ such that, given any $\eta_o > \eta_{\text{time}}$, one can find values $\tilde{\eta}_o > \eta_o$ and $\tilde{u}$ such that $\chi(\tilde{u}; \tilde{\eta}_o) \geq M$. First note that in order for $S_{\text{gen}}[\sigma(\tilde{\eta}_o), \Sigma(\tilde{\eta}_o)]$ to remain finite, it must be that the function $\chi(u; \tilde{\eta}_o)$ is only greater than M on an interval that has vanishing measure in the limit as $\tilde{\eta}_o \to \eta_\infty$. Otherwise, the proper area of $\sigma(\tilde{\eta}_o)$ diverges. Therefore, the leaves of $\mathcal{Q}$ develop "tendrils" in the asymptotic future limit, as illustrated in Fig. 6.7. In this case, however, the comoving volume that is enclosed by $\sigma(\tilde{\eta}_o)$

[9] A particularly astute reader may have noticed that the light cones in the example in Sec. 6.6 do not satisfy this global maximality property, but this is just because the approximation in which S_{out} is estimated by S_{CG} breaks down. More precisely, S_{CG} is not a good estimate of the matter contribution to generalized entropy for light cone slices that are far to the past of the light cone's tip. For such slices, the comoving volume enclosed by the slice grows arbitrarily large.

vanishes as $\tilde{\eta}_o \to \eta_\infty$, which means that the (locally) maximal entropy slice of each light cone coincides with the (globally) maximal area slice in the asymptotic future limit. We can then repeat the same arguments presented in the proof above for the constant-$\tilde{\eta}$ slice, but applied in comparison to the maximal area slice, to construct the required contradiction. Once it is established that $\mathcal{Q}$ squeezes into the comoving coordinate origin, the proof continues as before.

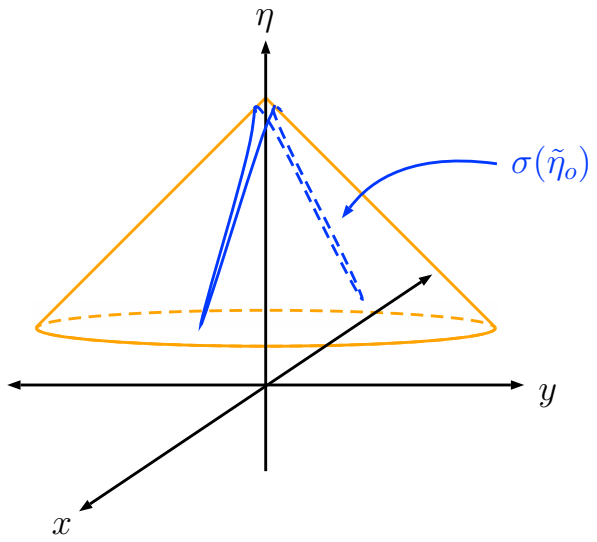

Figure 6.7: A hypothetical leaf $\sigma(\tilde{\eta}_o)$ that remains bounded away from the comoving coordinate origin. The leaf has two long tendrils that extend out from the comoving coordinate origin.

This relaxation is interesting (albeit somewhat artificial) because it makes it possible to avoid assuming the Quantum Focusing Conjecture. Moreover, as is shown in Sec. 6.7, if a RW spacetime admits a *continuous* holographic screen that has maximal area leaves on every past-directed light cone, then the screen itself is unique and there is always a finite globally maximal area slice of each light cone. (However, this slice is not necessarily unique and may not be part of the unique continuous holographic screen with leaves on every past-directed light cone.) This result suggests that it might in general be possible to relate continuity properties of screens to the properties of extremal-area light cone slices. For practical purposes, however, it is much cleaner to simply assume the QFC (which also guarantees that the GSL holds).

6.4.2 1+3 dimensions

Now suppose that $\mathcal{M}$ is a Bianchi spacetime in 1+3 dimensions with the line element

$$ds^2 = -dt^2 + a_1^2(t)\,dx^2 + a_2^2(t)\,dy^2 + a_3^2(t)\,dz^2\,. \tag{6.60}$$

The case where $\mathcal{M}$ has 3 dimensions of space parallels the 1+2-dimensional case with only a handful of technical complications. The main difference is that now the anisotropy has two functional degrees of freedom:

$$ds^2 = -dt^2 + a^2(t)\left[e^{2b_1(t)}dx^2 + e^{2b_2(t)}dy^2 + e^{2b_3(t)}dz^2\right]\,. \tag{6.61}$$

One arrives at the equation above by setting $a_i(t) = a(t)e^{b_i(t)}$ for $i = 1, 2, 3$, where the $b_i(t)$ are subject to the constraint $\sum_{i=1}^{3} b_i(t) = 0$. The definition of conformal light cone coordinates $(\eta, \eta_o, \theta, \phi)$ is correspondingly modified:

$$x^j(\eta, \eta_o, \theta, \phi) = D^j(\theta, \phi)\int_{\eta}^{\eta_o} \frac{e^{-2b_j(\zeta)}}{\sqrt{\sum_{i=1}^{3} D^i(\theta,\phi)^2\, e^{-2b_i(\zeta)}}}\,d\zeta, \tag{6.62}$$

where

$$D^j(\theta, \phi) = (\sin\theta\,\cos\phi, \sin\theta\,\sin\phi, \cos\theta).$$

Nevertheless, the essential construction remains unchanged. We still consider a past Q-screen, $\mathcal{Q}$, constructed with respect to a foliation of $\mathcal{M}$ by past-directed light cones, and the leaves of $\mathcal{Q}$ are still labeled by the conformal time η_o where the tip of their corresponding light cone is located. The no-hair theorem also generalizes in a straightforward way:

Theorem 6.4.4. *Let $\mathcal{M}$ be a Bianchi I spacetime with the line element* (6.60) *and whose matter content has constant thermodynamic entropy s per comoving volume. Suppose that $\mathcal{M}$ admits a past Q-screen, $\mathcal{Q}$, with globally maximal entropy leaves constructed with respect to a foliation of $\mathcal{M}$ with past-directed light cones that are centered on the origin, $x = y = z = 0$. Suppose that the Generalized Second Law holds on $\mathcal{Q}$ and that $\mathcal{M}$ and $\mathcal{Q}$ together satisfy the following assumptions for $i \in \{1, 2, 3\}$:*

(*i*) $a_i(t) \to \infty$ *as* $t \to \infty$,

(*ii*) $\mathcal{Q}$ *is timelike past some* $t_{\rm time}$ *and extends out to future timelike infinity,*

(*iii*) $\dot{a}_i(t) > 0$ *past some* $t_{\rm mono}$*,*

(*iv*) $S_{\rm gen} \to S_{\rm max} < \infty$ *along* $\mathcal{Q}$*.*

Then, $\mathcal{M}$ *is asymptotically de Sitter and the axial scale factors* $a_i(t)$ *approach* $C_i e^{Ht}$*, where* H *and* C_i *are constants.*

Note: In terms of $a(\eta)$ and the $b_i(\eta)$, Assumption (*i*) becomes:

(*i′*) $a(\eta) \to \infty$ as $\eta \to \eta_\infty$ and $a(\eta)e^{b_i(\eta)} \to \infty$.

Proof: The proof for 1+3 dimensions exactly parallels the proof of Theorem 6.4.1, so we only note the most important modifications. Beginning with Part 1, in $(\eta, \eta_o, \theta, \phi)$ coordinates, the leaves $\sigma(\eta_o)$ are now parameterized surfaces,

$$\tilde{x}^\mu(u, v; \eta_o) = (\eta(u, v; \eta_o), \eta_o, u, v) \qquad u \in [0, \pi] \qquad v \in [0, 2\pi)\,. \tag{6.63}$$

Our first task is again to show that $\chi(u, v; \eta_o) \equiv \eta_o - \eta(u, v; \eta_o)$ tends to zero for all values of u and v as $\eta_o \to \eta_\infty$.

As before, let us construct a contradiction of Assumption (*iv*) by supposing that $\mathcal{Q}$ never squeezes into the comoving coordinate origin. Suppose that there exists $M > 0$ such that, given any $\eta_o > \eta_{\rm time}$, one can find values $\tilde{\eta}_o > \eta_o$, $\tilde{u}$, and $\tilde{v}$ such that $\chi(\tilde{u}, \tilde{v}; \tilde{\eta}_o) \geq M$. Let $\tilde{\eta} \equiv \eta(\tilde{u}, \tilde{v}; \tilde{\eta}_o)$ and consider the constant $\eta = \tilde{\eta}$ slice of the light cone whose tip is at $\tilde{\eta}_o$. Denote this (co-dimension 2) surface by $\varsigma(\tilde{\eta}; \tilde{\eta}_o)$, and denote the (co-dimension 1) hypersurface of constant-$\tilde{\eta}$ by $X(\tilde{\eta})$. Here as well, Eq. (6.32) will lead us to the contradiction via a divergence in $A[\varsigma(\tilde{\eta}; \tilde{\eta}_o)]$.

In 1+3 dimensions, the induced metric on a surface of constant η and η_o is given by

$$\gamma_{ab} = \frac{\partial x^\mu}{\partial \theta^a} \frac{\partial x^\nu}{\partial \theta^b} g_{\mu\nu} = a^2(\eta) \sum_{j=1}^{3} e^{2b_j(\eta)} \frac{\partial x^j}{\partial \theta^a} \frac{\partial x^j}{\partial \theta^b}\,, \tag{6.64}$$

and where x^j and $g_{\mu\nu}$ refer to Eqs. (6.61) and (6.62) with $\theta^a \equiv (\theta, \phi)$. The area of this surface is now given by the surface integral

$$A(\eta, \eta_o) = \int_0^{\pi} \int_0^{2\pi} \sqrt{\gamma}\; d\phi\, d\theta\,, \tag{6.65}$$

where the determinant of the induced metric is

$$\gamma = a^4(\eta) \sum_{i<j} e^{2(b_i(\eta)+b_j(\eta))} \left(\frac{\partial x^i}{\partial \theta}\frac{\partial x^j}{\partial \phi} - \frac{\partial x^j}{\partial \theta}\frac{\partial x^i}{\partial \phi} \right)^2 . \tag{6.66}$$

One may therefore bound the area of $\varsigma(\eta; \eta_o)$ by, e.g.,

$$A(\eta, \eta_o) \geq a^2(\eta) e^{b_1(\eta)+b_2(\eta)} \iint d\theta\, d\phi\; \left| \frac{\partial x}{\partial \theta}\frac{\partial y}{\partial \phi} - \frac{\partial y}{\partial \theta}\frac{\partial x}{\partial \phi} \right| . \tag{6.67}$$

Using the coordinate transformation Eq. (6.62), one can show that the Jacobian in the integrand above is given by

$$\begin{aligned} \left| \frac{\partial x}{\partial \theta}\frac{\partial y}{\partial \phi} - \frac{\partial y}{\partial \theta}\frac{\partial x}{\partial \phi} \right| = \iint_{\eta}^{\eta_o} ds\, ds'\;\; & \sin\theta\, |\cos\theta| \left(\sin^2\theta \cos^2\phi\, e^{2(b_2(s)+b_3(s'))} \right. \\ & \left. + \sin^2\theta \sin^2\phi\, e^{2(b_3(s)+b_1(s'))} + \cos^2\theta\, e^{2(b_2(s)+b_1(s'))} \right) \\ & \times \left(\sum_{i=1}^{3} D^i(\theta,\phi)^2\, e^{-2b_i(s)} \right)^{-3/2} \left(\sum_{j=1}^{3} D^j(\theta,\phi)^2\, e^{-2b_i(s')} \right)^{-3/2} . \end{aligned} \tag{6.68}$$

This is quite beastly, but fortunately we can bound it nicely:

$$\begin{aligned} \left| \frac{\partial x}{\partial \theta}\frac{\partial y}{\partial \phi} - \frac{\partial y}{\partial \theta}\frac{\partial x}{\partial \phi} \right| \geq \iint_{\eta}^{\eta_o} ds\, ds'\;\; & \sin\theta \left|\cos^3\theta\right| e^{2(b_2(s)+b_3(s'))} \\ & \times \left((e^{-2b_1(s)} + e^{-2b_2(s)}) \sin^2\theta + e^{-2b_3(s)} \cos^2\theta \right)^{-3/2} \\ & \times \left((e^{-2b_1(s')} + e^{-2b_2(s')}) \sin^2\theta + e^{-2b_3(s')} \cos^2\theta \right)^{-3/2} . \end{aligned} \tag{6.69}$$

(One would arrive at similar results by choosing different terms to keep in the numerator of Eq. (6.68).) Then, inserting Eq. (6.69) into Eq. (6.67) and performing the angular integration, one arrives at

$$A(\eta, \eta_o) \geq 4\pi a^2(\eta) e^{b_1(\eta)+b_2(\eta)} \iint_{\eta}^{\eta_o} ds\, ds'\; \tilde{f}(s, s')\,,$$

where

$$\begin{aligned}\tilde{f}(s,s') &= \frac{e^{b_1(s)+3b_2(s)+3b_1(s')+b_2(s')}}{\left[\sqrt{e^{2b_1(s)}+e^{2b_2(s)}}e^{2(b_1(s')+b_2(s'))}+\sqrt{e^{2b_1(s')}+e^{2b_2(s')}}e^{2(b_1(s)+b_2(s))}\right]^2} \\ &\geq \frac{e^{b_1(s)+3b_2(s)+3b_1(s')+b_2(s')}}{\left[(e^{b_1(s)}+e^{b_2(s)})e^{2(b_1(s')+b_2(s'))}+(e^{b_1(s')}+e^{b_2(s')})e^{2(b_1(s)+b_2(s))}\right]^2} \\ &\equiv f(s,s')\,.\end{aligned}$$

Note that we have also used the fact that $b_3 = -b_1 - b_2$ to eliminate b_3. Then, if $f(s,s')$ is minimized at $s_m, s'_m \in [\eta, \eta_o]$, it follows that

$$A(\eta,\eta_o) \geq 4\pi(\eta_o-\eta)^2 a^2(\eta) e^{b_1(\eta)+b_2(\eta)} f(s_m,s'_m)\,. \tag{6.70}$$

Given this result, we now apply it to our surface $\varsigma(\tilde{\eta};\tilde{\eta}_o)$, for which $(\tilde{\eta}_o - \tilde{\eta}) \geq M$. Doing so, we arrive at

$$A[\varsigma(\tilde{\eta};\tilde{\eta}_o)] \equiv A(\tilde{\eta},\tilde{\eta}_o) \geq 4\pi M^2 a^2(\tilde{\eta}) e^{b_1(\tilde{\eta})+b_2(\tilde{\eta})} f(s_m,s'_m)\,. \tag{6.71}$$

The right-hand side of the bound above then diverges as $\tilde{\eta}$ and $\tilde{\eta}_o$ are chosen arbitrarily large. The only subtlety arises if either or both of b_1 and b_2 also diverge, but because the numerator and the denominator of the b-dependent part of the bound Eq. (6.71) contain equal powers of b_1 and b_2, the overall divergent behaviour induced by $a(\eta)$ is unchanged. (Recall that $a(\eta)e^{b_1(\eta)}$, $a(\eta)e^{b_2(\eta)}$, and $a(\eta)e^{b_3(\eta)} = a(\eta)e^{-b_1(\eta)-b_2(\eta)}$ all grow infinitely large by assumption.) We therefore arrive at the desired contradiction of Assumption (iv) via Eq. (6.32), and so the leaves of $\mathcal{Q}$ squeeze into the comoving coordinate origin in the asymptotic future.

Next we turn to showing that the scale factor $a(\eta)$ is asymptotically de Sitter (Part 2). Consider the generalized entropy $S_{\rm gen}[\sigma(\eta_o),\Sigma(\eta_o)]$ once more. First, Lemma 6.4.2 is correspondingly modified:

Lemma 6.4.5. *Let $\varsigma(\eta;\eta+\chi)$ be the constant-η slice of the past-directed light cone whose tip is at $\eta_o = \eta+\chi$. The generalized entropy defined by this slice is given by*

$$S_{\rm gen}[\varsigma(\eta;\eta+\chi),X(\eta)] = \frac{A(\eta,\eta+\chi)}{4G} + c_g(\eta,\chi)\chi^3 s, \tag{6.72}$$

where $A(\eta, \eta+\chi)$ is given by

$$A(\eta, \eta+\chi) = a^2(\eta) \cdot \left[4\pi\chi^2 + O(\chi^4)\right], \tag{6.73}$$

and $c_g(\eta, \chi)$ is some $O(1)$ geometric factor due to anisotropy that does not depend on $a(\eta)$.

Proof: Repeating the steps described in Lemma 6.4.2, one finds that

$$c_g(\eta, \chi) \equiv \frac{S_{\rm CG}(\eta, \chi)}{\chi^3 s} = \frac{4\pi}{3} + \frac{8\pi}{45}\left(b_1'(\eta)^2 + b_1'(\eta) b_2'(\eta) + b_2'(\eta)^2\right)\chi^2 + O(\chi^3). \tag{6.74}$$

The expansion of $A(\eta, \eta+\chi)$ for small χ follows from expanding $\sqrt{\gamma}$ in Eq. (6.65) in powers of χ and then integrating. ■

From Lemma 6.4.5, it therefore again follows that the matter contribution to the generalized entropy, $S_{\rm CG}[\sigma(\eta_o), \Sigma(\eta_o)]$, vanishes in the asymptotic future limit. Consequently, we focus on the area term, $A[\sigma(\eta_o)]$.

For this part of the proof, we will work in the coordinates $(\chi, \eta_o, \theta, \phi)$. The leaf $\sigma(\eta_o)$ is parameterized by some surface $\tilde{x}^\mu(u, v) = (\chi(u, v; \eta_o), \eta_o, u, v)$ with η_o held constant and $0 \leq u \leq \pi$, $0 \leq v < 2\pi$. In the asymptotic future, this surface is the surface on the light cone with tip at η_o with maximal area, and so it is the solution of

$$\frac{\delta A[\sigma(\eta_o)]}{\delta \chi(u, v; \eta_o)} = 0\,. \tag{6.75}$$

The induced metric on this surface is, as usual, given by

$$h_{ab} = \frac{\partial \tilde{x}^\mu}{\partial u^a}\frac{\partial \tilde{x}^\nu}{\partial u^b}\tilde{g}_{\mu\nu}\,, \tag{6.76}$$

where $\tilde{g}_{\mu\nu}$ is the metric of Eq. (6.61) but rewritten in $(\chi, \eta_o, \theta, \phi)$ coordinates. The area of $\sigma(\eta_o)$ is given by

$$A[\sigma(\eta_o)] = \int_0^\pi \int_0^{2\pi} \sqrt{\det h}\; dv\, du\,, \tag{6.77}$$

and the components of h_{ab} are as follows:

$$\begin{aligned} h_{uu} &= (\partial_u\chi)^2\tilde{g}_{00} + 2(\partial_u\chi)\tilde{g}_{02} + \tilde{g}_{22} \\ h_{uv} &= (\partial_u\chi)(\partial_v\chi)\tilde{g}_{00} + (\partial_u\chi)\tilde{g}_{03} + (\partial_v\chi)\tilde{g}_{02} + \tilde{g}_{23} \\ h_{vv} &= (\partial_v\chi)^2\tilde{g}_{00} + 2(\partial_v\chi)\tilde{g}_{03} + \tilde{g}_{33}. \end{aligned} \tag{6.78}$$

Once more, solving the full Euler-Lagrange problem for $\chi(u, v; \eta_o)$ to obtain the maximal area A is intractable, so we use the same trick where we extract an overall factor of $a^4(\eta_o - \chi)$ from $\det h$ and then expand the square root of the quotient in powers of χ. The result is

$$A[\sigma(\eta_o)] = \int_0^{\pi}\int_0^{2\pi} a^2(\eta_o - \chi)\left[\frac{\sin\theta}{R(u,v;\eta_o)^{3/2}}\chi^2 + \frac{Q(u,v;\eta_o)\sin\theta}{R(u,v;\eta_o)^{5/2}}\chi^3 + O(\chi^4)\right] dv\, du, \tag{6.79}$$

where

$$\begin{aligned} R(u,v;\eta_o) &= \sum_{i=1}^{3} e^{-2b_i(\eta_o)} D^i(u,v)^2 \\ Q(u,v;\eta_o) &= \sum_{i=1}^{3} b_i'(\eta_o) e^{-2b_i(\eta_o)} D^i(u,v)^2. \end{aligned} \tag{6.80}$$

Only keeping the lowest order term, the variation $\delta A/\delta\chi = 0$ gives the maximal path $\chi(u, v; \eta_o) = \chi(\eta_o)$ as the solution of

$$\chi = \frac{a(\eta_o - \chi)}{a'(\eta_o - \chi)}. \tag{6.81}$$

So, to lowest order, $A[\sigma(\eta_o)]$ is given by

$$A[\sigma(\eta_o)] = \chi^2 a^2(\eta_o - \chi)\int_0^{\pi}\int_0^{2\pi} \frac{\sin\theta}{R(u,v;\eta_o)^{3/2}}\, dv\, du = 4\pi\left(\frac{a^2(\eta_o - \chi)}{a'(\eta_o - \chi)}\right)^2. \tag{6.82}$$

But the requirement that $S_{\text{gen}} \to S_{\max}$ means that $A[\sigma(\eta_o)]/4G$ must tend to the constant value $S_{\max}$, or in other words,

$$\lim_{\substack{\eta_o \to \eta_\infty \\ \chi \to 0}} \frac{a^2(\eta_o - \chi)}{a'(\eta_o - \chi)} = \sqrt{\frac{GS_{\max}}{\pi}} \equiv \frac{1}{H}. \tag{6.83}$$

Therefore, $a(\eta)$ asymptotically approaches de Sitter, $a(\eta) \to -1/H\eta$ as $\eta \to 0^-$, with $H^2 = \pi/GS_{\max}$. Note that we recover the same Hubble constant as in Theorem 6.3.1 for RW spacetimes in 1+3 dimensions.

Finally, as in the case of 1+2 dimensions, the condition $(a(\eta)e^{b_i(\eta)})' > 0$ is enough to show that $\lim_{\eta\to 0^-} b_i'(\eta)$ exists for each i. □

6.5 Discussion

Assuming the Generalized Second Law, we have shown that if a Bianchi I spacetime admits a past Q-screen along which generalized entropy increases up to a finite maximum, then this implies that the spacetime is asymptotically de Sitter. We recover a version of Wald's cosmic no-hair theorem by making thermodynamic arguments about spacetime, without appealing to Einstein's equations.

While the proof of these cosmic no-hair theorems is most tractable (and certainly easiest to visualize) in 1+2 dimensions, the generalization to 1+3 dimensions was fairly immediate. In principle, the proof strategy for arbitrary dimensions is the same, albeit more difficult from the perspective of calculation. This is chiefly because calculating area elements of codimension-2 surfaces in arbitrary dimensions is cumbersome. Nevertheless, it is natural to expect that analogous cosmic no-hair theorems hold for Bianchi I spacetimes of arbitrary dimensions.

Within the proof itself, it would be interesting to see if the monotonicity assumptions, $a_i'(\eta) > 0$, could be eliminated. The fact that the Generalized Second Law asserts that S_{gen} increases monotonically along a Q-screen does offer some leverage. In particular, asymptotically this implies that the average scale factor

$$a(\eta) = \left(\prod_{i=1}^{d} a_i(\eta)\right)^{1/d}$$

increases monotonically; however, we learn nothing about the anisotropies $b_i(\eta)$, since the leading order behaviour of S_{gen} does not depend on the $b_i(\eta)$ in the asymptotic future regime. We also note that the monotonicity assumptions do not trivialize the cosmic no-hair theorems demonstrated in Sec. 6.4. For example, assuming monotonicity does not rule out exponential expansion with different rates in different spatial directions, nor asymptotically power-law scale factors, nor does it even imply accelerated expansion at all.

An interesting extension would be to try to prove a no-hair theorem for classical cosmological perturbations [191], or for quantum fields in curved spacetime. Given a scalar field on a curved spacetime background, the task is to show that the combined metric and scalar field perturbations approach the Bunch-Davies state [192] at late times. In principle it would suffice to show that the background spacetime still tends to de Sitter in the future in this case, since one could then simply invoke known no-hair results about scalar fields in curved backgrounds [129–131]. Conceptually, such a calculation would be interesting because one can explicitly write down the quantum state of cosmological perturbations, and so a full treatment of the matter entropy as von Neumann entropy (modulo ultraviolet divergences) is possible.

To prove our theorem, it was not strictly necessary to assume that the gravitational contribution to the entropy was precisely proportional to the surface area. We could imagine choosing some other function of the area, such that

$$S_{\rm gen}[\sigma, \Sigma] = f(A[\sigma]/G) + S_{\rm CG}[\sigma, \Sigma]. \tag{6.84}$$

For example, returning to the RW case, if one sets $f(A/G) = C(A/G)^p$ for some constants C and p, exactly the same analysis as in the proof of Theorem 6.3.1 leads to the conclusion that (cf. Eq. (6.17))

$$\dot{a}(t) \to \sqrt{\frac{4\pi}{G}\left(\frac{C}{S_{\rm max}}\right)^{1/p}}\, a(t) \tag{6.85}$$

in the limit as $t \to \infty$. In other words, one still concludes that the scale factor is asymptotically de Sitter, albeit with a Hubble constant that differs from the usual case of $f(A/G) = A/4G$.

Finally, while we did not make use of the Einstein field equations in our derivation, upon reinvoking them, we note that the cosmic no-hair theorems established here imply a pure dark energy phase asymptotically in the future (in the sense that the stress energy tensor becomes proportional to the metric, $g_{\mu\nu}$). However, the GSL is not sensitive to the nature of the dark energy (whether it is a pure cosmological constant, whether it turns on, whether it's due to a slowing scalar field, and so on).

This work can be thought of as part of the more general program of connecting gravitation to entropy, thermodynamics, and entanglement [20, 24, 65–67, 76, 86, 93, 170–174]. As in attempts to derive Einstein gravity from entropic considerations, we deduce the behavior of the geometry of spacetime from thermodynamics, without explicit field equations. Our result is less general, as we only obtain the asymptotic behavior of the universe, but is perhaps also more robust, as our assumptions are correspondingly minimal. Thinking of spacetime as emerging thermodynamically from a set of underlying degrees of freedom can change our perspective on the knotty problems of quantum gravity; for example, as emphasized by Banks [137], the cosmological constant problem becomes the question of "Why does Hilbert space have a certain number of dimensions?" rather than "Why is this parameter in the low-energy effective Lagrangian so small?" Problems certainly remain (including why the entropy was so low near the Big Bang), but this alternative way of thinking about gravitation may prove useful going forward.

Acknowledgements

We would like to thank Cliff Cheung, John Preskill, and Alan Weinstein for helpful discussions. This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of High Energy Physics, under Award No. DE-SC0011632, as well as by the Walter Burke Institute for Theoretical Physics at Caltech and the Gordon and Betty Moore Foundation through Grant No. 776 to the Caltech Moore Center for Theoretical Cosmology and Physics.

6.6 Q-screens, a worked example

In this section, we illustrate Q-screens by explicitly constructing one in a RW spacetime that is asymptotically de Sitter. Consider a RW spacetime in 1+3 dimensions with the line element $ds^2 = -dt^2 + a^2(t)(d\chi^2 + \chi^2 d\Omega_2^2)$ and where the scale factor is $a(t) = \sinh t$, $t \in (0, \infty)$. Conformal time is given by $\eta(t) = -2\,\mathrm{arccoth}(e^t)$,

$\eta \in (-\infty, 0)$, and the scale factor in conformal time is

$$a(\eta) = \frac{1}{\sinh(-\eta)} \,. \tag{6.86}$$

Foliate the spacetime with past-directed light cones centered at the coordinate origin $\chi = 0$, and let the Cauchy hypersurfaces of the spacetime be the constant-η hypersurfaces. Let us now construct a Q-screen by extremizing the generalized entropy on each light cone.

Consider a past-directed light cone whose tip is at the conformal time η_o. A constant-$\eta < \eta_o$ slice of this light cone is a 2-sphere of coordinate radius $\eta_o - \eta$, and so the generalized entropy computed with respect to this slice is

$$S_{\text{gen}}(\eta; \eta_o) = \frac{\pi}{G} \left(\frac{\eta_o - \eta}{\sinh(-\eta)} \right)^2 + \frac{4}{3} \pi (\eta_o - \eta)^3 s \,. \tag{6.87}$$

A plot of $S_{\text{gen}}(\eta; \eta_o)$ as a function of η for several values of η_o is shown in Fig. 6.8. The area term $A(\eta; \eta_o)/4G$ alone is also overlaid on the plot, which illustrates that it is the dominant contribution to the generalized entropy at late times. Notice that in addition to having a local maximum, $S_{\text{gen}}(\eta; \eta_o)$ also has a local minimum, and below a certain critical value η_o^{crit} there is in fact no nonzero value of η which locally extremizes $S_{\text{gen}}(\eta; \eta_o)$. As such, the Q-screen, which is defined as the union of the slices with maximal generalized entropy, is only defined for $\eta_o \geq \eta_o^{\text{crit}}$. This is in contrast to the area $A(\eta; \eta_o)$, which has a locally maximizing value of η for all η_o. The holographic screen, which is made up of extremal *area* slices, is therefore defined for all times. Both the Q-screen and the holographic screen were schematically illustrated previously in Fig. 6.2.

Generalized entropy is extremal when $\partial S_{\text{gen}}/\partial \eta = 0$. Excluding $\eta = 0$ and $\eta \to -\infty$, the extremizing values of η are the real-valued solutions of

$$\eta_o - \eta = \frac{\sinh(-\eta)}{\cosh(-\eta) - 2Gs \sinh(-\eta)^3} \tag{6.88}$$

when they exist. Let $\eta^Q(\eta_o)$ denote the maximizing value, and hence also define the Q-screen leaf radius $\chi^Q(\eta_o) \equiv \eta_o - \eta^Q(\eta_o)$. A plot of $\chi^Q(\eta_o)$ is shown in Fig. 6.9. As expected, $\chi^Q(\eta_o)$ vanishes as $\eta_o \to 0^-$. For comparison, we also plot the holographic

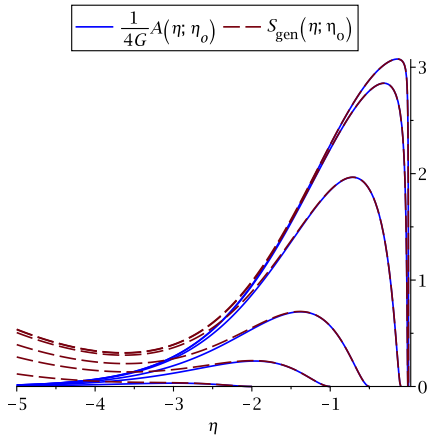

Figure 6.8: Plots of area (solid) and generalized entropy (dashed) along light cones. From the lowest peak to the highest peak, the values of η_o are -2, -1, -0.5, -0.1, -0.01, and -0.001. Here we have taken $G = 1$ and we have picked $s = 0.001$.

screen radius $\chi^H(\eta_o) \equiv \eta_o - \eta^H(\eta_o)$, where $\eta^H(\eta_o)$ maximizes the area of the light cone slice, i.e., it is the solution of

$$\eta_o - \eta = \tanh(-\eta)\,. \tag{6.89}$$

In particular note that $\chi^Q(\eta_o)$ is always slightly larger than $\chi^H(\eta_o)$, but they ultimately coincide in the limit $\eta \to 0^-$ (cf. Fig. 6.2).

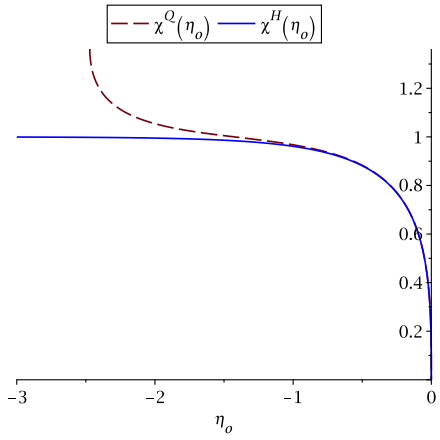

Figure 6.9: Asymptotic behaviour of the radius of the Q-screen leaves ($\chi^Q(\eta_o)$, dashed) and holographic screen leaves ($\chi^H(\eta_o)$, solid).

As a final exercise, let us investigate the asymptotic dependence of $\chi^H(\eta_o)$ on η_o (which is also the asymptotic dependence of $\chi^Q(\eta_o)$, since the two coincide as $\eta_o \to 0^-$) to illustrate some of the subtleties involved in performing asymptotic expansions.

Consider Eq. (6.89) and let $\eta = \eta_o - \chi$ so that we have $\chi = \tanh(\chi - \eta_o)$. Since, asymptotically, $\chi \to 0$, one may be tempted to expand this last equation for small values of χ:

$$\chi = \tanh(-\eta_o) + (1 - \tanh^2(-\eta_o))\chi + O(\chi^2) \quad \Rightarrow \quad \chi^H(\eta_o) \stackrel{?}{=} \frac{1}{\tanh(-\eta_o)} \tag{6.90}$$

Notice, however, that since $0 < \tanh(-\eta_o) < 1$, this expression for $\chi^H(\eta_o)$ cannot be infinitesimally small—the expansion is inconsistent! Rather, χ and η_o are simultaneously infinitesimal. Consider instead the double Taylor series in χ and η_o:

$$\chi = \chi - \eta_o - \tfrac{1}{3}\chi^3 + \eta_o\chi^2 - \eta_o^2\chi + \tfrac{1}{3}\eta_o^3 + \cdots \quad \Rightarrow \quad \chi^H(\eta_o) = (-3\eta_o)^{1/3} + \eta_o + \cdots \tag{6.91}$$

This last result is the correct asymptotic behaviour of $\chi^H(\eta_o)$.

Similarly, writing $A = 4\pi\chi^2 a^2(\eta_o - \chi)$, one arrives at the wrong expressions for extremal values if one tries to expand A in small values of χ, η_o, or even both at the same time. The key is to keep $a(\eta_o - \chi)$ intact so that one arrives at Eq. (6.89). Doing so leaves just enough nonlinearity to be able to restore the correct asymptotic behaviour of $\chi^H(\eta_o)$. This technique is exploited in Sec. 6.4.1.

6.7 Holographic screen continuity and maximal area light cone slices

When the null curvature condition holds, the Raychaudhuri equation guarantees that light rays focus, or in other words, that the expansion of a null congruence is always nonincreasing: $d\theta/d\lambda \leq 0$. In particular, this means that if a null congruence has a spacelike slice whose area is maximal with respect to local deformations, then this is in fact the unique globally maximal area slice. A consequence of this observation is that if one's aim is to construct a holographic screen by stitching together maximal area slices of each null sheet in a null foliation, then the holographic screen is uniquely fixed by the choice of foliation.

Here, we connect the uniqueness of locally maximal area slices to continuity properties of holographic screens in RW spacetimes. What we will first show is that, given a foliation of a RW spacetime by past-directed light cones, there is at most one continuous holographic screen that can be constructed with respect to this foliation

that has maximal area leaves on every light cone. We will then show that a consequence of this observation is that if a spacetime admits a continuous holographic with maximal area leaves on every light cone, then each light cone necessarily has a globally maximal finite area slice.

Proposition 6.7.1. *Let $\mathcal{M}$ be a RW spacetime with the line element*

$$ds^2 = a^2(\eta)\left(-d\eta^2 + d\chi^2 + \chi^2 d\Omega_{d-1}^2\right), \tag{6.92}$$

where the conformal time η takes values in an unbounded (connected) interval $\mathcal{I} \subseteq \mathbb{R}$. Consider a foliation of $\mathcal{M}$ by past-directed light cones whose tips are at $\chi = 0$. If there is a past-directed light cone that has multiple spacelike slices that have maximal area with respect to local deformations, then $\mathcal{M}$ admits at most one holographic screen, H, constructed with respect to the given foliation that is both (a) continuous, and (b) has maximal area leaves on every past-directed light cone.

Proof: Consider a past-directed light cone whose tip is at η_o. For $\eta < \eta_o$, the area of the constant-η slice of this light cone is given by

$$A(\eta, \eta_o) = \mathcal{N}_d \left[(\eta_o - \eta)a(\eta)\right]^{d-1}, \tag{6.93}$$

where $\mathcal{N}_d$ is a dimension-dependent constant. Because $\mathcal{M}$ is spherically symmetric, such a slice has extremal area if $\partial A/\partial\eta = 0$, or equivalently, if

$$\eta_o = \eta + \frac{a(\eta)}{a'(\eta)} \equiv f(\eta). \tag{6.94}$$

Therefore, constant-η slices of the past-directed light cone whose tip is at η_o for which $f(\eta) = \eta_o$ and $\eta < \eta_o$ are potential holographic screen leaves.

Now suppose that there is a light cone whose tip is at η_o that has n locally maximal area slices at $\eta = \eta_1, \eta_2, ..., \eta_n$ where, for convenience, these conformal times are ordered such that $\eta_1 > \eta_2 > \cdots > \eta_n$. This means that a graph of $f(\eta)$ must intersect the horizontal line at η_o at least $2n - 1$ times. (Between any two adjacent local maxima η_i and η_{i+1}, there must be a local minimum of area at some $\eta_i^{\min} \in (\eta_i, \eta_{i+1})$, and there may also be inflection points.) Consider any two adjacent local maxima

η_i and η_{i+1}. Schematically, in the vicinity of these points, the graph of $f(\eta)$ must look like one of the two configurations shown in Fig. 6.10 (a,b), since f is continuous if a/a' is continuous. Now consider shifting the horizontal line at η_o up and down. This corresponds to shifting the tip of the light cone to the future and past of η_o. Where the horizontal line intersects the graph of $f(\eta)$ tracks how the location of the local maxima and the local minimum of A move. In particular, notice that by moving the horizontal line sufficiently far to the future or the past, one of the local maxima and the local minimum must eventually meet and become an inflection point before disappearing altogether. (Note that we may always move the horizontal line sufficiently far in at least one of the past or future directions, since the interval $\mathcal{I}$ in which the conformal time takes its values is unbounded in at least one direction.) Therefore, if we track how the locations of the maxima at η_i and η_{i+1} change as we move the location of the light cone's tip, we see that one of these local maxima must eventually disappear, as illustrated in Fig. 6.10 (d).

Inductively, then, there exists at most one *continuous* function, call it $\eta_{\max}(\eta_o)$, whose domain is all $\eta_o \in \mathcal{I}$ and is such that $\eta = \eta_{\max}(\eta_o)$ is a local maximum of $A(\eta, \eta_o)$ for all η_o. The union of the constant-$\eta_{\max}(\eta_o)$ slices of all past-directed light cones is precisely the holographic screen H described in the statement of the proposition. □

Examples of various $f(\eta)$ are sketched below. Fig. 6.11 (a) depicts a case in which there exists a continuous holographic screen with leaves on every light cone. Fig. 6.11 (b) depicts a case in which there is no such holographic screen. In fact, from this example, one can see that if $\mathcal{I} = \mathbb{R}$, then there can never be a continuous holographic screen with leaves on every light cone if there is a light cone that has multiple maximal area slices. Referring to the proof above, the technical reason is that in this case, the horizontal line of constant η_o can be pushed arbitrarily far up and down since η_o can take all values in $\mathbb{R}$, and so any pair of adjacent maxima and minima will eventually merge (as a function of η_o).

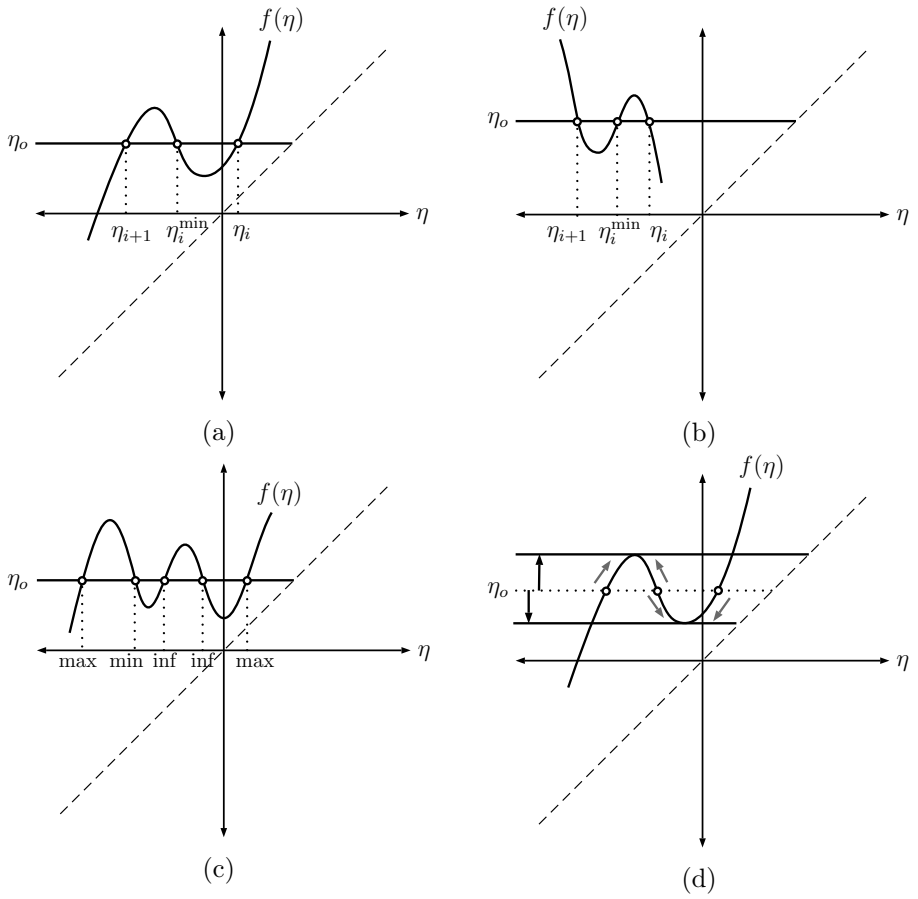

Figure 6.10: In the vicinity of two local maxima of $A(\eta, \eta_o)$ at η_i and η_{i+1} with η_o held fixed, the graph of the function $f(\eta)$ must schematically resemble one of the configurations shown in (a) and (b). The dashed line is the $\eta_o = \eta$ line; η can only take values to the left of this line. The graph (c) depicts a configuration with additional inflection points. The graph (d) illustrates how the maxima and the minimum eventually meet and annihilate as the horizontal constant η_o line is shifted up and down.

Finally, there is a partial converse of the result above:

Proposition 6.7.2. *If $\mathcal{M}$ as described in Proposition 6.7.1 admits a continuous holographic screen, H, with maximal area leaves on every past-directed light cone, then each light cone has a spatial slice which is a global maximum of the area of all spatial slices of the light cone, and the area of this slice is finite.*

Proof: Consider first the case where there is a unique local maximum on each past-directed light cone and H is the union of these maximal area surfaces. Again denote the value of η that maximizes $A(\eta, \eta_o)$ for a given η_o by $\eta_{\max}(\eta_o)$. The only way that

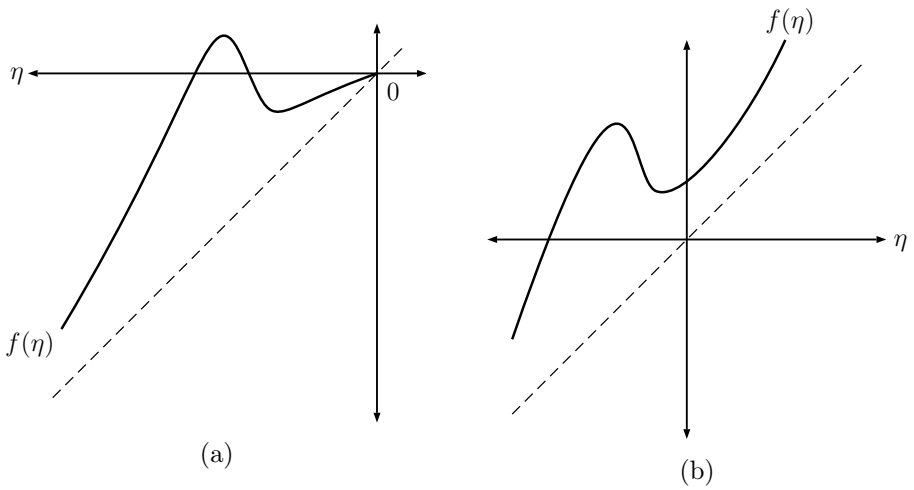

Figure 6.11: (a) An example of a function $f(\eta)$, where $\mathcal{I} = (-\infty, 0)$, which admits a continuous holographic screen with leaves on every light cone, but where some light cones have two local maxima of $A(\eta, \eta_o)$ when their tips are close enough to the endpoint $\eta_o = 0$. (b) An example of a function $f(\eta)$, where $\mathcal{I} = \mathbb{R}$, which admits no continuous holographic screen with leaves on every light cone.

$\eta_{\max}(\eta_o)$ could *not* be a global maximum of area is if there was some $\eta_M < \eta_{\max}(\eta_o)$ such that $A(\eta_M, \eta_o) > A(\eta_{\max}(\eta_o), \eta_o)$. However, for this to be possible, there must be a local minimum of A in between η_M and $\eta_{\max}(\eta_o)$. In other words, the function $f(\eta)$ must intersect the horizontal constant-η_o line once for the local maximum, once for the local minimum, and then possibly an additional even number of times for pairs of inflection points—there cannot be more intersections if $\eta_{\max}(\eta_o)$ is the unique local maximum of area. This means that the graph of $f(\eta)$ must be concave up or concave down (Fig. 6.12), in which case there will be some horizontal η_o lines that do not intersect the graph of $f(\eta)$, which contradicts the requirement that H have leaves on every light cone. Therefore, $\eta = \eta_{\max}(\eta_o)$ is in fact a global maximum of $A(\eta, \eta_o)$.

Then, according to Proposition 6.7.1, the other case is where some light cones have multiple local maxima of $A(\eta, \eta_o)$, in which case $\mathcal{I}$ is only semi-infinite. This can only happen for light cones whose tips are near the finite endpoint of the interval $\mathcal{I}$. Beyond some threshold value of η in the direction in which $\mathcal{I}$ is unbounded, $f(\eta)$ must still be monotonic in order for there to be leaves on every light cone. Therefore, for η_o beyond the threshold, the first case applies, and when there are multiple local

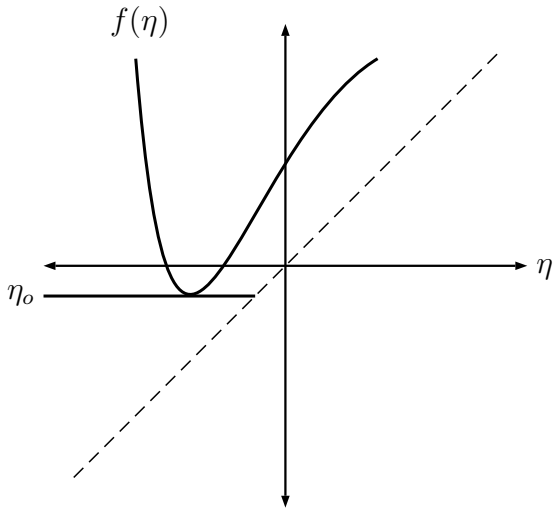

Figure 6.12: An example of $f(\eta)$ that is concave up, with $\mathcal{I} = \mathbb{R}$. Above the horizontal η_o line shown, this $f(\eta)$ intersects a horizontal line twice, which means that light cones that correspond to such horizontal lines have a local maximum and a local minimum of $A(\eta, \eta_o)$. However, light cones that correspond to horizontal lines drawn below the horizontal line shown have no local extrema, since these lines do not intersect $f(\eta)$.

maxima of area on a given light cone for η_o between the threshold and the finite endpoint of $\mathcal{I}$, at least one of them is a global maximum of $A(\eta, \eta_o)$. $\square$

Chapter 7

NATURAL COVARIANT PLANCK SCALE CUTOFFS AND THE COSMIC MICROWAVE BACKGROUND

We calculate the impact of quantum gravity–motivated ultraviolet cutoffs on inflationary predictions for the cosmic microwave background spectrum. We model the ultraviolet cutoffs fully covariantly to avoid possible artifacts of covariance breaking. Imposing these covariant cutoffs results in the production of small, characteristically $k-$dependent oscillations in the spectrum. The size of the effect scales linearly with the ratio of the Planck to Hubble lengths during inflation. Consequently, the relative size of the effect could be as large as one part in 10^5, i.e., eventual observability may not be ruled out.

This chapter was published as Ref. [6], A. Chatwin-Davies, A. Kempf, and R. T. W. Martin, "Natural covariant Planck scale cutoffs and the cosmic microwave background spectrum," Phys. Rev. Lett. ***119*** *(2017) 031301,* arXiv:1612.06445.

7.1 Introduction

It is widely expected that the very notion of distance in space and time breaks down at or before the Planck scale, due to quantum fluctuations of the metric (see [193] for a review). To understand the structure of spacetime at the Planck scale will, therefore, require a theory of quantum gravity. There are several current approaches to a consistent theory of quantum gravity, including string theory, loop quantum gravity, and others [194, 195]. It has proven exceedingly difficult, however, to test models for Planck-scale physics experimentally, chiefly due to the extremely small scales involved.

One of the most promising approaches to experimentally probing quantum gravity theories is to look for small imprints that Planck-scale physics may have left in the cosmic microwave background (CMB) and the subsequent structure formation

[196–201]. This is because, according to the standard inflationary scenario, the quantum fluctuations that seeded the CMB's inhomogeneities likely originated only about 5 or 6 orders of magnitude away from the Planck scale [201], namely when the comoving modes' fluctuations froze at their Hubble horizon crossing.

The magnitude of the imprint of Planck-scale physics in the CMB has been estimated using various approaches. These generally model the influence that Planck-scale physics exerts on inflationary quantum field theory through an ultraviolet (UV) cutoff at the Planck scale and through generalized dispersion relations [200,202–213]. Of crucial importance is the question of how the magnitude of the effect scales with the ratio of the Planck and Hubble lengths during inflation [201,214]. Let us denote the Hubble scale at the end of inflation by L_{Hubble} and define $\sigma = L_{\mathrm{Planck}}/L_{\mathrm{Hubble}}$. Several studies found that the imprint of quantum gravity in the CMB should be of the order of σ^{α} with either $\alpha = 1$ [206] or $\alpha = 2$ [200]. If indeed $\alpha = 1$, then we may be a mere five orders of magnitude away from measuring Planck-scale physics, which is relatively close when compared to accelerator physics, where there are about 15 orders of magnitude to cross.

Models for how Planck-scale physics influences inflationary quantum field theory, such as modified dispersion relations, generally break local Lorentz invariance [201]. This makes it unclear to what extent the predicted imprints on the CMB are due to Planck-scale physics and to what extent the predicted imprints are caused by the breaking of local Lorentz invariance. Therefore, to isolate the effect of Planck-scale physics on the CMB, we employ functional analytic methods that allow us to model natural UV cutoffs fully covariantly.

Our main results are that covariant UV cutoffs can produce small characteristic oscillations in the fluctuation spectrum and that this imprint on the CMB is of first order, i.e., $\alpha = 1$. The overall amplitude is sensitive to the timing of the comoving modes' quantum-to-classical transition. This transition is generally expected to have occurred soon after horizon crossing [22, 215–218]. This is because, after horizon crossing, comoving modes quickly become extremely squeezed, which makes them

extremely susceptible to environment-induced decoherence [219]. We will be able to conclude, therefore, that the relative size of the imprint of a covariant Planck-scale cutoff in the CMB could be as large as one part in 10^5 so that eventual observability may not be ruled out.

7.2 Fully covariant natural ultraviolet cutoffs.

Our aim is to covariantly model possible first order corrections to standard quantum field theory (QFT) as the Planck scale is approached from low energies. To this end, recall that in the path integral formulation of QFT, the path integral is normally assumed to run over all field configurations. This includes on-shell fields, i.e., fields that extremize the action, as well as fields that are arbitrarily far off-shell. In order to ensure the preservation of covariance, we will here consider UV cutoffs that remove or suppress field configurations—and therefore field fluctuations—that are off-shell by an amount that is on the order of or past the Planck scale. Technically, the eigenfunctions of the d'Alembertian to eigenvalues beyond the Planck scale will be considered too far off-shell and will therefore be modeled as being suppressed or eliminated entirely from the path integration. Such cutoffs are manifestly covariant and diffeomorphism invariant since the spectra of covariant differential operators such as the d'Alembertian, and the corresponding operators for fields other than scalars, are independent of the choice of coordinates.

Concretely, let $\Box$ denote a self-adjoint d'Alembertian on a Lorentzian manifold, $\mathcal{M}$. Its eigenfunctions form an orthonormal basis for $L^2(\mathcal{M})$ and can be taken to span the space of functions in the quantum field theoretic path integral. A cutoff on the spectrum of $\Box$ is operationally defined via (real linear combinations of) spectral projectors,

$$f(\Box) = \sum_{\lambda \in \mathrm{spec}(\Box)} f(\lambda)\, \langle \psi_\lambda, \cdot \rangle\, \psi_\lambda \,, \tag{7.1}$$

where f is the non-negative function which defines the cutoff, ψ_λ is an eigenfunction of $\Box$, and $\langle \cdot, \cdot \rangle$ denotes the inner product on $L^2(\mathcal{M})$.

Choosing $f(\lambda) = \theta(\Omega^2 - |\lambda|)$ produces a sharp cutoff on the spectrum of $\Box$ at $|\lambda| = \Omega^2$.

We will focus on this most extreme case of an UV cutoff. It is straightforward to smooth out the step function so that the spectral cutoff is not as sharp and the positive operator $f(\Box)$ describes models of smoother UV cutoffs. These covariant cutoffs amount to suppressing far off-shell field fluctuations from the path integral.[1]

The class of cutoffs that we consider is the most general type of kinematic covariant cutoff within the framework of QFT. We are not going beyond the framework of QFT because inflation appears to have happened well within the range of validity of QFT. Among these cutoffs, we focus on the most extreme case, the sharp cutoff, to obtain a prediction for the maximal impact on the fluctuation spectrum.

As an aside, we also note that with the above choice of a sharp cutoff f, the quantity Ω can be interpreted as a covariant bandlimit: a conventional bandlimit, i.e., a minimum wavelength imposed on functions in $\mathbb{R}^n$, can be thought of as a cutoff on the spectrum of the Laplacian, $\triangle$, on $\mathbb{R}^n$. The eigenfunctions of $\triangle$ are the plane waves $\exp(i\mathbf{k}\cdot\mathbf{x})$ and cutting off the spectrum of $\triangle$ is to impose a limit on the length of the wave vector k^2. As a consequence, Shannon's sampling theorem applies: any function that is Ω-bandlimited is determined everywhere if known on any regular discrete lattice $\{x_n\}$ with a spacing smaller than $1/(2\Omega)$ [220–222], in each direction. The d'Alembertian generalizes the Laplacian to Lorentzian spacetimes, leading to a covariant generalization of sampling theory [223, 224]: Each mode of a particular spatial wavelength possesses a corresponding bandwidth and therefore obeys a sampling theorem in time. Those modes whose wavelengths are smaller than the Planck length possess an exceedingly small bandwidth, which effectively freezes them out. This ultraviolet behavior is beautifully covariant, as the notions of spatial wavelength and temporal bandwidth Lorentz transform appropriately.

Returning to our main program, we express the space of covariantly-bandlimited scalar fields on $\mathcal{M}$ in terms of eigenfunctions of the d'Alembertian as

$$B_{\mathcal{M}}(\Omega) = \text{span}\left\{\psi \mid \Box\psi = \lambda\psi, \lambda \in [-\Omega^2, \Omega^2]\right\}. \tag{7.2}$$

[1] We remark that, since proton decay is mediated by far off-shell processes, such a covariant ultraviolet cutoff may help explain the exceedingly large proton lifetime.

Here, Ω sets the ultraviolet scale. In general, the spectrum of a self-adjoint d'Alembert operator is not bounded below for Lorentzian-signature manifolds, and so the spectrum must be cut off from above and below.

In the path integral formulation of QFT, we implement the covariant bandlimitation by only integrating over covariantly-bandlimited fields instead of all field configurations. For example, the covariantly-bandlimited Feynman propagator of a quantized scalar field, which we denote by G_F^Ω, is given by

$$iG_F^\Omega(x,x') = \frac{\int_{B_{\mathcal{M}}(\Omega)} \mathcal{D}\phi \; \phi(x)\phi(x') e^{iS[\phi]}}{\int_{B_{\mathcal{M}}(\Omega)} \mathcal{D}\phi \; e^{iS[\phi]}} \,. \tag{7.3}$$

Covariant bandlimitation here amounts to excluding the most extreme off-shell fluctuations from the quantum field theoretic path integral. Concretely, G_F^Ω can be computed by acting on the conventional propagator G_F to the left and right with the spectral projectors $\theta(\Omega^2 - \Box) \equiv P_\Omega$, where $G_F(x,x')$ is understood to be the kernel of an integral operator.

7.3 Application to inflation.

Let $\mathcal{M}$ be an inflating Friedmann-Robertson-Walker (FRW) spacetime with the line element $ds^2 = a^2(\eta)[-d\eta^2 + d\mathbf{x}^2]$, where the conformal time η takes values in an interval $\mathcal{I} \subseteq (-\infty, 0)$. Consider a massless scalar field ϕ on this background. Such a field is a proxy for quantities such as the Mukhanov-Sasaki variable, which describes combined quantized perturbations of the inflaton and scalar metric degrees of freedom, or tensor perturbations of the metric, which describe primordial gravitational waves.

The strength of the field's quantum fluctuations is quantified by its fluctuation spectrum:

$$\begin{aligned} \delta\phi_k(\eta) &= \tfrac{1}{2\pi} k^{3/2} |v_k(\eta)| \\ &= \sqrt{4\pi} k^{3/2} |G_F(\eta = \eta'; k)|^{1/2} \,. \end{aligned} \tag{7.4}$$

Here, $v_k(\eta)$ is the field's mode function and k is comoving wavelength. We define the covariantly-bandlimited fluctuation spectrum by replacing G_F in the equation

above with $G_F^\Omega = P_\Omega G_F P_\Omega$. Note that the spectrum of the d'Alembertian in a FRW spacetime is preserved (up to degeneracy) under spatial Fourier transforms, i.e., if $\Box u(\eta, \mathbf{x}) = \lambda u(\eta, \mathbf{x})$, then $\Box_k u_\mathbf{k}(\eta) = \lambda u_\mathbf{k}(\eta)$. This allows us to calculate the bandlimited propagator comoving mode by comoving mode.

We now consider the flat slicing of de Sitter spacetime in 1+3 dimensions, which is the most computationally tractable model for an inflating FRW spacetime. The scale factor is $a(\eta) = -1/H\eta$ with $-\infty < \eta < 0$, and H is the Hubble parameter. We take the state of the field to be the Bunch-Davies vacuum [192] so that the Feynman propagator without cutoff reads, as usual:

$$G_F(\eta, \eta'; k) = -\frac{i\pi}{4} \frac{\sqrt{\eta\eta'}}{a(\eta)a(\eta')} \Big[\theta(\eta - \eta') H^{(1)}_{3/2}(k|\eta|) H^{(2)}_{3/2}(k|\eta'|) + \theta(\eta' - \eta) H^{(2)}_{3/2}(k|\eta|) H^{(1)}_{3/2}(k|\eta'|)\Big]. \tag{7.5}$$

Here, $H^{(1)}_{3/2}$ and $H^{(2)}_{3/2}$ denote Hankel functions of the first and second kind, respectively.

Our strategy is as follows: for each comoving mode k, construct the spectral projectors P_Ω from the eigenfunctions and eigenvalues of $\Box_k$, and then apply these to the left and right of $G_F(\eta, \eta'; k)$ to obtain G_F^Ω. Equivalently, we can write $P_\Omega = I - P_\Omega^\perp$, where $P_\Omega^\perp$ projects onto eigenspaces corresponding to $|\lambda| > \Omega^2$, so that

$$G_F^\Omega = G_F - (P_\Omega^\perp G_F + G_F P_\Omega^\perp - P_\Omega^\perp G_F P_\Omega^\perp) \tag{7.6}$$

and the quantity in brackets gives the correction to the full propagator.

We notice first that each k-d'Alembertian on $L^2((-\infty, 0), a^4(\eta)d\eta)$ is not uniquely self-adjoint. In functional analytic language [225–227], the minimal symmetric operator generated by each $\Box_k$ has deficiency indices $(1, 1)$, which implies the existence of a one-parameter family of self-adjoint extensions of $\Box_k$, each corresponding to a generalized boundary condition. We identify the correct self-adjoint extension by matching the generalized boundary condition to the boundary condition implied by

G_F as given in Eq. (7.5). The fact that G_F is a right inverse of the d'Alembertian, i.e., that $\Box_k G_F(\eta, \eta'; k) = a^{-4}(\eta)\delta(\eta - \eta')$, means that G_F is diagonal in the same basis as $\Box_k$ and so it shares the same boundary condition.

For each $\lambda \in \mathbb{R}$, the eigenfunction equation $\Box_k u(\eta) = \lambda u(\eta)$ yields a Sturm-Liouville differential equation,

$$(a^2 u')' + k^2 a^2 u + \lambda a^4 u = 0. \tag{7.7}$$

This can be solved to obtain two linearly independent solutions [225]. For $\lambda < 9H^2/4$, Eq. (7.7) admits one normalizable solution and so the self-adjoint extensions of $\Box_k$ will have point spectrum in this range. For $\lambda \geq 9H^2/4$, both solutions are non-normalizable, and so this range can only contain continuous spectrum. The possible normalizable solutions for $\lambda < 9H^2/4$ are

$$\psi_n(\eta) = H^2 \sqrt{2p_n}\, |\eta|^{3/2} J_{p_n}(k|\eta|)\,, \tag{7.8}$$

where $p_n = p_0 + 2n$ with $n \in \mathbb{N}$ ensures orthonormality and the value of $p_0 \in (0, 2]$ fixes the self-adjoint extension. (Here, J_p denotes the Bessel-J function of order p). The corresponding eigenvalues are $\lambda_n = H^2(\frac{9}{4} - p_n^2)$. We determined the correct choice of self-adjoint extension by examining the action of $G_F^h - \lambda_n^{-1} I$ on test eigenfunctions as a function of p_0, where G_F^h denotes the Hermitian part of G_F. When p_0 takes the value that is implied by G_F as given in Eq. (7.5), then $(G_F^h - \lambda_n^{-1} I)\psi_n(\eta\,; p_0)$ must be in the kernel of $\Box_k$. By varying p_0 and checking when this last condition is satisfied, we found that $p_0 = 3/2$ (so that $\lambda_0 = 0$) is the self-adjoint extension that is implied by Eq. (7.5).

Orthonormality then implies that all $\lambda \geq 9H^2/4$ are in the continuous spectrum, and that the corresponding eigenfunctions are

$$\begin{aligned}\psi_{q(\lambda)}(\eta) =& H^2 \sqrt{\tfrac{1}{2} q \tanh(\pi q)}\, \Big[\operatorname{sech}(\tfrac{\pi}{2} q) |\eta|^{3/2} \operatorname{Re} J_{iq}(k|\eta|) \\ & - \operatorname{csch}(\tfrac{\pi}{2} q) |\eta|^{3/2} \operatorname{Im} J_{iq}(k|\eta|) \Big]\,,\end{aligned} \tag{7.9}$$

with $q(\lambda) = (\frac{\lambda}{H^2} - \frac{9}{4})^{1/2}$. The eigenfunctions of the continuous spectrum have been

normalized so that

$$\int_{-\infty}^{0} \psi_q(\eta)\psi_{q'}(\eta)\ a^4(\eta)d\eta = \delta(q-q')\,. \tag{7.10}$$

We deduced this normalization numerically by requiring that

$$\int_{q-\epsilon}^{q+\epsilon}\int_{-\infty}^{0} \psi_q(\eta)\psi_{q'}(\eta)\ a^4(\eta)d\eta = 1. \tag{7.11}$$

Upon exchanging the order of integration, the double integral becomes Riemann-integrable and thus calculable numerically.

Assembling the results, the projectors $P^{\perp}_{\Omega}$ are given by

$$P^{\perp}_{\Omega}(\eta,\eta') = \sum_{n>N} \psi_n(\eta)\psi_n(\eta') + \int_{Q}^{\infty} \psi_q(\eta)\psi_q(\eta')\ dq\,, \tag{7.12}$$

where $Q = q(\Omega^2)$ and $N = \max\{n : |\lambda_n| < \Omega^2\}$. The full correction to the propagator in Eq. (7.6) can then be calculated using a combination of exact antiderivatives when possible and numerical integration otherwise. The contribution from the point spectrum is several tens of orders of magnitude smaller than the contribution from the continuous spectrum, and is therefore negligible in Eq. (7.12). A plot of $\Delta(\delta\phi_k)/\delta\phi_k$ is shown in Fig. 7.1, where

$$\begin{aligned}\Delta(\delta\phi_k(\eta)) =& \sqrt{4\pi}k^{3/2}\left(|G^{\Omega}_F(\eta=\eta';k)|^{1/2}\right.\\ &\left.-|G_F(\eta=\eta';k)|^{1/2}\right).\end{aligned} \tag{7.13}$$

The quantity $\Delta(\delta\phi_k)/\delta\phi_k$ characterizes the magnitude of the impact of the covariant UV cutoff on inflationary predictions for the CMB.

Fig. 7.1 shows that the impact of the covariant bandlimit on inflationary perturbations is sensitive to when the perturbations were in effect measured and became classical. As is well known, after horizon crossing, comoving modes quickly become highly squeezed. This squeezing makes the quantum statistics of field fluctuations indistinguishable from those of a classical stochastic ensemble, which leads to an "apparent" quantum-to-classical transition [217]. If the comoving modes had continued to evolve as a closed system until re-heating, then our calculations, which

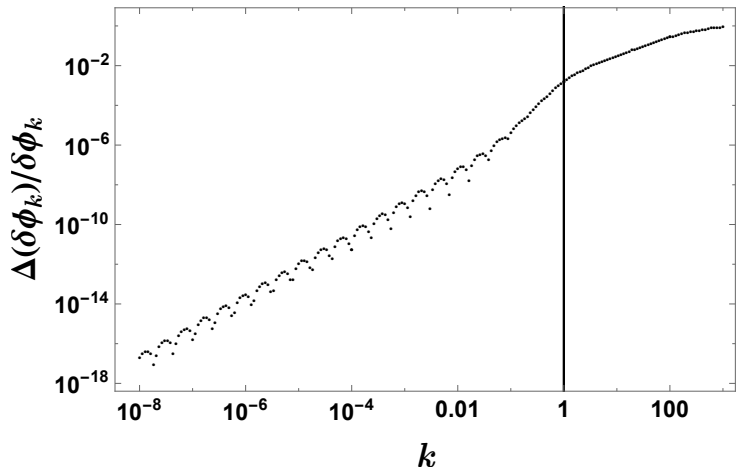

Figure 7.1: Relative change in $\delta\phi_k$ as a function of comoving mode number with $|\eta| = 1$ fixed. Horizon crossing ($k|\eta| = 1$) is marked with a solid line. The ratio of Planck to Hubble scales is set to $H/\Omega = 10^{-2}$ so that the numerical computation remains tractable. The oscillations are due to an oscillatory integrand that appears in G_F^{Ω} which dominates at small k. Roughly, the oscillations may be understood as an interference effect, with dips occurring when an integer number of wavelengths fit between the Planck and Hubble scales.

hold as long as the modes are independent closed systems, would be valid up until re-heating. At this point the re-heating interactions will decohere the inflationary perturbations, which would then be described according to open system dynamics. In this scenario, the imprint that Planck-scale physics could have left in the CMB would, in effect, be measured and fixed as late as re-heating. The imprint would therefore be exponentially suppressed.

However, the modes' squeezing at horizon crossing is generally expected to have also led to a quantum-to-classical transition through environmental decoherence already shortly after horizon crossing, therefore requiring an open system description at that stage. For our calculation, this means that the imprint of Planck-scale physics in the CMB was in effect measured soon after horizon crossing. We therefore predict that this imprint is not suppressed.

For completeness, let us briefly discuss why decoherence is expected to have occurred soon after horizon crossing (for a more extensive review, see [217]). During inflation, there are several environmental sources of decoherence for the comoving modes.

These include nonlinear gravitational interactions among the modes as well as interactions with the fields of other species. (There is also the effect of decoherence from tracing over degrees of freedom beyond the cosmological horizon.) The weakness of these environmental interactions is offset by the well-known [219] extreme sensitivity of highly-squeezed states, such as the comoving modes after horizon crossing, to environmental decoherence. We remark that these decoherence mechanisms are indeed such that the pointer basis is approximately the field eigenbasis and that the modes' standing wave behavior can account for the acoustic oscillations in the CMB, as required by phenomenology [217].

Within this standard scenario for the quantum-to-classical transition in inflation, we can then conclude that the cutoff-induced modulation of the primordial quantum fluctuations' amplitudes that we calculate was effectively fixed by measurement through environmental decoherence near horizon crossing. The imprint in the CMB scales as σ^α with $\alpha = 1$, and here, $\sigma = L_{\text{Planck}}/L_{\text{Hubble}} \sim H/\Omega$ could be as large as 10^{-5} for realistic values of H and Ω. While still far from being measurable, it is conceivable that such an effect might eventually become observable, thereby providing some access to Planck-scale physics.

We determined the order, α, by holding $k|\eta|$ fixed and plotting $\Delta(\delta\phi_k)/\delta\phi_k$ as a function of H/Ω. This is shown in Fig. 7.2, where the scaling behavior can be read off: the effect scales almost exactly linearly, i.e., $\alpha \approx 1$, when $k|\eta|$ is fixed close to horizon crossing.

We now consider the realistic case of inflationary spacetimes with a slowly-varying Hubble parameter. The exact calculations would be challenging since the corresponding d'Alembertian would be computationally even more difficult to diagonalize. We therefore model this case with an "adiabatic" approximation in which the spacetime is instantaneously de Sitter at every conformal time η, but in which we let H slowly vary as a function of η. Fig. 7.3 shows a plot of $\Delta(\delta\phi_k)/\delta\phi_k$ for this model. At each k, $|\eta|$ is set to $1/k$ and the Hubble parameter is set to the value taken at the mode's horizon crossing by the time-varying Hubble parameter $H(\eta)$ of a power

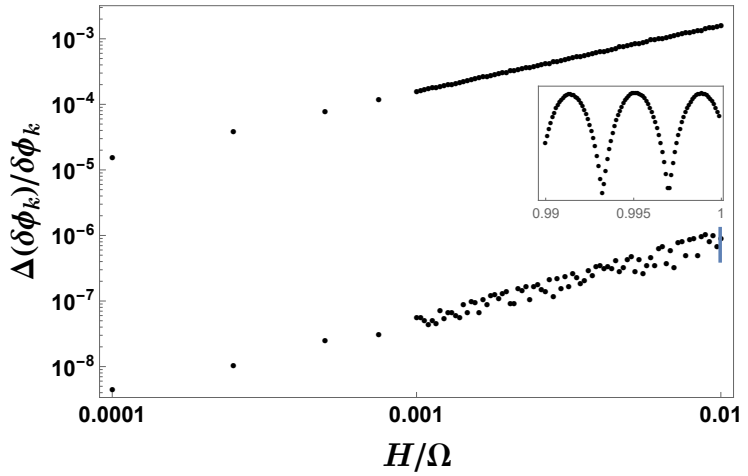

Figure 7.2: Relative change in $\delta\phi_k$ as a function of the ratio of Planck to Hubble scales. The upper series of points is at horizon crossing ($k|\eta| = 1$), and the lower series is well past horizon crossing ($k|\eta| = 1/20$). The inset shows a sliver of the lower series at higher resolution, where the apparent scatter in the data is resolved as rapid oscillations.

law spacetime. As the magnitude of $\Delta(\delta\phi_k)/\delta\phi_k$ tracks the effective time-varying Hubble parameter, an interesting characteristic pattern of oscillations appears. Intuitively, these oscillations may be thought of as arising from the time-varying number of Planckian wavelengths that fit into a Hubble length. If observed, such oscillations in the CMB spectra may serve as an experimental signature of a covariant natural UV cutoff that could not easily be alternatively explained through a plausible inflaton potential. Our results are consistent with prior literature [199, 204, 205] in that we also predict superimposed oscillations. However, our new predictions for the type and magnitude of such oscillations are obtained covariantly and our predictions, including the prediction that the effect is of first order ($\alpha = 1$), are now free of potential artifacts due to covariance breaking.

Acknowledgements

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of High Energy Physics, under Award Number DE-SC0011632, as well as by the Walter Burke Institute for Theoretical Physics at Caltech and the Gordon and Betty Moore Foundation through Grant 776 to the Caltech Moore

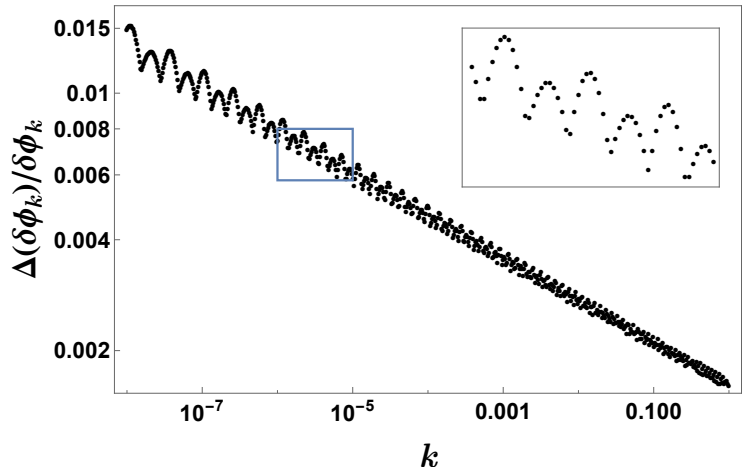

Figure 7.3: Relative change in the fluctuation spectrum for a slowly-varying Hubble parameter. For each k, $|\eta|$ is set to $1/k$ and the Hubble parameter is matched to that of a power law spacetime: $H(\eta) = p[C(p-1)|\eta|]^{1/(p-1)}$, where $a(t) = Ct^p$ in cosmic time. Here we set $p = 10$ and we fixed C such that $H(-1) = 1$, so as to coincide with horizon crossing in Fig. 7.1. The value of H ranges from 1 to 7.74 as k goes from 1 to 10^{-8}; this change is tracked by the linear trend in $\Delta(\delta\phi_k)/\delta\phi_k$. The inset shows a detail of the oscillations on top of the trend.

Center for Theoretical Cosmology and Physics. AK and ACD acknowledge support from the Natural Sciences and Engineering Research Council of Canada (NSERC). RTWM acknowledges support by the National Research Foundation (NRF) of South Africa, CPRR grant 90551.

Part III

The Black Hole Information Problem

Chapter 8

HAWKING RADIATION AND INFORMATION: A REVIEW

An important discovery in theoretical physics during the last century was Stephen Hawking's finding that black holes radiate, which, over timescales of order M^3 for a black hole of initial mass M, should cause a black hole to dissipate into a dilute ensemble of radiation. The topic of Hawking radiation is a point of overlap between gravity and quantum theory, and the black hole information problem is an especially puzzling open question. In its modern form, the essential implication of the black hole information problem is that black holes either do not evolve unitarily, or our understanding of local effective field theory and black holes as quantum mechanical objects must change when gravity is taken into account.

Since an appreciable portion of this thesis concerns the black hole information problem, this chapter reviews the phenomenon of Hawking radiation as well as the black hole information problem. The topics covered, in the order that they follow, are radiation from black holes that formed from collapsing matter, thermal properties of eternal black holes, and finally the black hole information problem in both its original form (as addressed by Hawking [43], Page [40, 228], and Susskind *et. al* [23]), as well as its modern form (due to Almheiri *et. al* [18, 19]).

My motivation for writing this chapter was to create a chance for myself to carefully go through Hawking's original work and the accompanying original literature. As a result, the derivation in Sec. 8.1 of the production of radiation from black holes that formed from collapsing matter is quite detailed. The discussion about thermal properties of eternal black holes in Sec. 8.2 is less detailed, but has all of the necessary ingredients for setting up the black hole information problem in Sec. 8.3. In this last section, I attempt to be as careful and precise as I can in my formulation of the problem. In particular, the argument given for the black hole information problem in terms of strong subadditivity of von Neumann entropy is the argument that I

believe to be most robust, and I make an effort to identify where assumptions and approximations are made in the argument.

This chapter draws heavily on Hawking's original article [43], as well as the texts by Wald [160], Birrell & Davies [192], and Mukhanov [229]. Other specific references are indicated throughout when drawn upon.

8.1 Black holes that formed from collapse

Consider a black hole that formed from the spherical collapse of matter of total mass M.[1] That is, we shall consider a spacetime $\mathcal{M}$ such that, outside of some timelike surface $r(t)$ which describes the surface of the collapsing matter, the metric is that of a Schwarzschild black hole in the usual coordinates (t, r, θ, ϕ):

$$ds^2 = -\left(1 - \frac{2M}{r}\right) dt^2 + \left(1 - \frac{2M}{r}\right)^{-1} dr^2 + r^2 d\theta^2 + r^2 \sin^2\theta \, d\phi^2 \tag{8.1}$$

The relevant Penrose diagram is shown in Fig. 8.1.

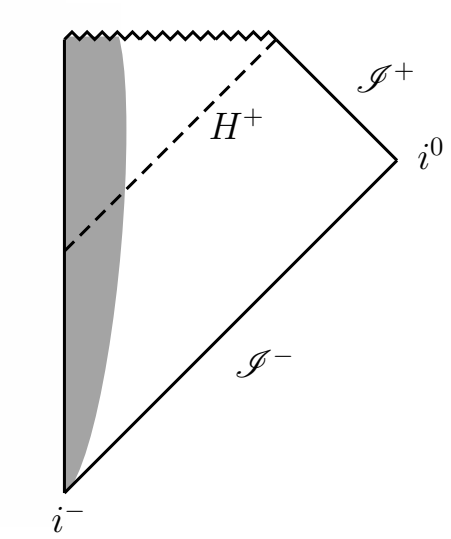

Figure 8.1: The Penrose diagram for a black hole that formed from the collapse of matter. The collapsing matter is shown as the grey region. The black hole horizon, H^+, is indicated by the dashed null line. The singularity at $r = 0$ is indicated by the jagged line.

Let φ be a scalar field of mass μ on this spacetime. With a view of eventually discussing the quantized field's mode functions, let us first investigate the classical

[1] To be certain, Hawking radiation is still produced in situations that are not spherically symmetric. For some discussion, refer to Ref. [43].

propagation of the field, i.e., let us investigate the solutions of its classical equation of motion, the Klein-Gordon equation:

$$(\Box - \mu^2)\varphi = 0\,. \tag{8.2}$$

As usual, the d'Alembertian is given by

$$\Box = \frac{1}{\sqrt{-g}}\partial_\alpha\left(\sqrt{-g}g^{\alpha\beta}\partial_\beta\right). \tag{8.3}$$

Since $\mathcal{M}$ is spherically symmetric, it makes sense to expand φ in terms of spherical harmonics,

$$\varphi(t,r,\theta,\phi) = \frac{1}{r}f(t,r)Y_l^m(\theta,\phi). \tag{8.4}$$

Using this ansatz, one finds using Eq. (8.2) and the metric (8.1) that, outside of the collapsing matter where $\mathcal{M}$ is Schwarzschild, the function $f(t,r)$ obeys

$$-\frac{\partial^2 f}{\partial t^2} + \frac{\partial^2 f}{\partial r_*^2} - \left(1-\frac{2M}{r}\right)\left(\frac{l(l+1)}{r^2} + \frac{2M}{r^3} + \mu^2\right) f = 0, \tag{8.5}$$

where

$$r_* = r + 2M\ln\left|\frac{r}{2M} - 1\right| \tag{8.6}$$

is the usual tortoise coordinate. Therefore, in the coordinates (t, r_*), we can think of $f(t,r_*)$ as a wave that scatters in a potential

$$V(r) = \left(1-\frac{2M}{r}\right)\left(\frac{l(l+1)}{r^2} + \frac{2M}{r^3} + \mu^2\right), \tag{8.7}$$

at least in the region of $\mathcal{M}$ that lies outside the collapsing matter.

Consider now the asymptotic behaviour of $V(r)$. When $r_* \to -\infty$ (or equivalently $r \to 2M$), for all values of the field's mass, one finds that

$$V(r) \sim 1 - \frac{2M}{r} \sim e^{r_\star/2M}\,. \tag{8.8}$$

The potential decays exponentially quickly in r_*, so solutions $f(t,r_*)$ will behave like free waves $e^{-i\omega(t\pm r_*)}$ near the black hole's event horizon. In the other asymptotic limit as $r_* \to \infty$, we have that

$$V(r) \sim \begin{cases} \left(1-\frac{2M}{r}\right)\mu^2 & \mu > 0 \\ \frac{l(l+1)}{r^2} & \mu = 0 \end{cases}. \tag{8.9}$$

Therefore, in the massless case, the solutions $f(t, r_*)$ will also behave like asymptotically free waves near the future and past asymptotic boundaries of $\mathcal{M}$, but the massive case receives a correction due to the $1/r$ potential.

For simplicity, from this point onward we will assume that $\mu = 0$ and only study the massless case. A heuristic way to motivate this assumption is by observing that the effective potential for massive fields is larger than the massive case. Since Hawking radiation consists of particle content that reaches the asymptotic future of $\mathcal{M}$, one should expect that a higher effective potential would suppress the emission of massive Hawking radiation and that massless species will dominate the radiation. This expectation is indeed confirmed by careful numerical calculations [230, 231]. It is still possible to perform an analysis for massive fields that is similar to what follows, but we refer the reader to Ref. [192] for details. Similarly, while we will limit ourselves to scalar fields, the analysis also generalizes to fields with nonzero spin [43, 192].

We now turn to the quantization of φ. A comprehensive review of the quantization of fields in curved spacetimes may be found in Ref. [192]. Alternatively, the second chapter of Ref. [232] gives a brief introduction to the quantization of a scalar field in a curved spacetime that is very much congruous with the development here. Very briefly, though, recall that the quantized field φ in a globally hyperbolic spacetime may be faithfully written as[2]

$$\varphi(x) = \sum_i \left(f_i(x) a_i + \bar{f}_i(x) a_i^\dagger \right), \tag{8.10}$$

where $a_i^\dagger$ and a_i are creation and annihilation operators that obey

$$[a_i, a_j^\dagger] = \delta_{ij}. \tag{8.11}$$

The functions $f_i(x)$ are solutions of the equation of motion (8.2) and are called *mode functions*. That the mode functions solve the Klein-Gordon equation ensures that

[2]The discrete sum should only be taken as a schematic illustration; in certain cases (such as those we will encounter below), the sums may be continuous integrations, in which case the Kronecker delta function normalization converts to Dirac delta function normalization.

the field operator itself obeys the Klein-Gordon equation. Furthermore, the mode functions must be chosen so that

$$(f_i, f_j) \equiv \frac{i}{2} \int_\Sigma \left(f_i \bar{f}_{j;a} - \bar{f}_j f_{i;a} \right) d\Sigma^a = \delta_{ij} \tag{8.12}$$

for any given Cauchy surface Σ. One can show that if Eq. (8.12), also called the *Wronskian condition*, holds for one Cauchy surface, then it holds on all Cauchy surfaces. The Wronskian condition guarantees that the field's canonical commutation relations hold everywhere on the manifold. The mode functions then also form a complete orthonormal basis with respect to the bilinear form (8.12).

The mode functions are solutions of the Klein-Gordon equation (8.2), so they may be fixed by specifying initial data on a Cauchy surface. Since our field is massless, two natural Cauchy surfaces to consider in light of the discussion above are the past null boundary of spacetime, $\mathscr{I}^-$, and the union of the black hole event horizon and future null boundary of spacetime, $H^+ \cup \mathscr{I}^+$.

First consider the past null boundary. That solutions of the Klein-Gordon equation behave as free waves here motivates choosing the mode functions to be

$$f_{\omega lm}(t, r, \theta, \phi) = \frac{1}{\sqrt{2\pi\omega}} \frac{1}{r} F_{\omega l}(r) e^{-i\omega v} Y_l^m(\theta, \phi), \tag{8.13}$$

where $v = t + r_*$. The function $F_{\omega l}(r)$ is fixed by the Klein-Gordon equation and the Wronskian condition, but must approach the value 1 as $r \to \infty$ as dictated by the asymptotic structure of solutions of Eq. (8.2). These mode functions $f_{\omega lm}$ therefore describe free inward-propagating spherical waves at $\mathscr{I}^-$ with frequency ω (as measured with respect to the null time v)[3] and angular quantum numbers l and m. The full field then takes the form

$$\varphi = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \int_0^\infty d\omega \ \left(f_{\omega lm} a_{\omega lm} + \bar{f}_{\omega lm} a^\dagger_{\omega lm} \right) . \tag{8.14}$$

With this choice of mode functions, the operators $a^\dagger_{\omega lm}$ and $a_{\omega lm}$ may be naturally interpreted as creating and annihilating quanta of ingoing free spherical waves at

[3]Note that this choice of positive frequency modes is opposite to Hawking's original convention, i.e., Hawking's mode functions are proportional to $e^{i\omega v}$ in Ref. [43]. The convention used here is the same as in Ref. [160] and also happens to be the convention that I usually use.

$\mathscr{I}^-$. The vacuum state defined by these operators, i.e., the state $|0^-\rangle$ such that

$$a_{\omega lm}|0^-\rangle = 0 \quad \forall\ \omega, l, m \tag{8.15}$$

therefore corresponds to the state with no incoming excitations of the field in the asymptotic past.

Now consider the future boundary $H^+ \cup \mathscr{I}^+$. That the future boundary is bipartite suggests defining new mode functions that are split into two sets, $\{p_{\omega lm}\}$ and $\{q_{\omega lm}\}$, where the $p_{\omega lm}$ vanish on the event horizon and the $q_{\omega lm}$ vanish on $\mathscr{I}^+$. Here we choose

$$p_{\omega lm}(t, r, \theta, \phi) = \frac{1}{\sqrt{2\pi\omega}} \frac{1}{r} P_{\omega l}(r) e^{-i\omega u} Y_l^m(\theta, \phi), \tag{8.16}$$

where $u = t - r_*$, so that the mode functions $p_{\omega lm}$ describe outward-propagating, free spherical waves at $\mathscr{I}^+$. With a similar choice for the $q_{\omega lm}$, we can conclude that these mode functions describe free waves that are crossing into the black hole at its event horizon. Then, together with a set of new creation and annihilation operators, $b_{\omega lm}, b^\dagger_{\omega lm}$ and $c_{\omega lm}, c^\dagger_{\omega lm}$, we can expand the field operator in terms of these new modes:

$$\varphi = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \int_0^\infty d\omega \ \left(p_{\omega lm} b_{\omega lm} + \bar{p}_{\omega lm} b^\dagger_{\omega lm} + q_{\omega lm} c_{\omega lm} + \bar{q}_{\omega lm} c^\dagger_{\omega lm} \right). \tag{8.17}$$

We briefly note that there is some ambiguity in how to choose the modes $q_{\omega lm}$, which is discussed in more detail in Ref. [160]. However, as long as we are only concerned with making predictions about quantities at $\mathscr{I}^+$, then the choice of the $q_{\omega lm}$ does not matter for the following reason. In reference to the abstract mode decomposition (8.10), the structure of the resulting Fock space is naturally a tensor product over the modes i:

$$\mathcal{H} = \bigotimes_i \left(\bigoplus_{n_i} \mathrm{span}\{|n_i\rangle\} \right). \tag{8.18}$$

In the equation above, $|n_i\rangle$ denotes the state obtained from n_i applications of the creation operator $a_i^\dagger$ to the vacuums state. In particular, this means that we can arrange the Hilbert space, defined with respect to the Cauchy surface $H^+ \cup \mathscr{I}^+$, as

a tensor product of Fock states for the $b_i, b_i^\dagger$ and $c_i, c_i^\dagger$:

$$\mathcal{H}_{H^+ \cup \mathscr{I}^+} = \mathcal{H}_{H^+} \otimes \mathcal{H}_{\mathscr{I}^+} \,. \tag{8.19}$$

In other words, we can just group those indices i in Eq. (8.18) that correspond to the $p_{\omega lm}$ mode functions into $\mathcal{H}_{\mathscr{I}^+}$, and group those indices that correspond to the $q_{\omega lm}$ into $\mathcal{H}_{H^+}$. Then, provided that we are only interested in making predictions at $\mathscr{I}^+$, we can always trace out the part of the state that has support on $\mathcal{H}_{H^+}$. The resulting reduced state on $\mathcal{H}_{\mathscr{I}^+}$ is unaffected by redefinitions of the mode functions $q_{\omega lm}$.

Since the first set of mode functions, $\{f_{\omega lm}, \bar{f}_{\omega lm}\}$ forms a complete basis, we can express the new mode functions $p_{\omega lm}$ as a linear combination of the former for some coefficients $\alpha_{\omega\omega'}$ and $\beta_{\omega\omega'}$:

$$p_{\omega lm} = \int_0^\infty d\omega' \ \left(\alpha_{\omega\omega'} f_{\omega' lm} + \beta_{\omega\omega'} \bar{f}_{\omega' l(-m)}\right) \,. \tag{8.20}$$

This expansion, via the commutation relations of ϕ and $a_{\omega lm}$, $a^\dagger_{\omega lm}$ imply that

$$b_{\omega lm} = \int_0^\infty d\omega' \ \left(\bar{\alpha}_{\omega\omega'} a_{\omega' lm} - \bar{\beta}_{\omega\omega'} a^\dagger_{\omega' l(-m)}\right) . \tag{8.21}$$

In particular, we will find that the state $|0^-\rangle$ is not annihilated by the $b_{\omega lm}$. In other words, starting with no field excitations in the asymptotic past, we will find excitations on $\mathscr{I}^+$ with nonzero probability:

$$\langle 0^- | b^\dagger_{\omega lm} b_{\omega lm} | 0^- \rangle = \int_0^\infty d\omega' \ |\beta_{\omega\omega'}|^2 \,. \tag{8.22}$$

This is precisely the Hawking radiation.

The program to calculate the spectrum of the produced radiation is therefore quite simple in principle:

- Write down $f_{\omega lm}$.
- Write down $p_{\omega lm}$.
- Use the bilinear form (8.12) to extract $(p_{\omega lm}, \bar{f}_{\omega' l(-m)}) = \beta_{\omega\omega'}$.

However, the functions $F_{\omega l}(r)$ and $P_{\omega l}(r)$ are not expressible in terms of elementary functions, and so the best we can do without resorting to numerics is an approximate calculation, such as the one that follows.

Here we will follow the basic approach outlined in Ref. [160] with some hybridization with Hawking's original work [43]. Our strategy will be to consider a solution $p_{\omega lm}$ starting at $\mathscr{I}^+$ and study how it propagates back to $\mathscr{I}^-$ so that we may compute its overlap with $f_{\omega lm}$.

First, note that the plane wave-like part of these mode functions, $e^{-i\omega u}$, causes the mode function's oscillation frequency to diverge near the event horizon. Indeed, any freely-falling observer who crosses into the black hole would report that these mode functions have a diverging oscillation frequency, which we can confirm by rewriting $e^{-i\omega u}$ in terms of the observer's geodesic in a coordinate chart that is smooth across the horizon.

Namely, recall that the null Kruskal-Szekeres coordinate U is related to u by

$$U = -e^{-\kappa u}, \tag{8.23}$$

where $\kappa = 1/(4M)$ is the surface gravity of a Schwarzschild black hole, and that the Kruskal-Szekeres coordinate system smoothly covers the black hole horizon with the horizon located at $U = 0$ (refer to Eqs. (8.41-8.43) and Fig. 8.5 below). Therefore, in terms of U, we may write

$$e^{-i\omega u} = e^{-i\omega\kappa^{-1}\ln(-U)}. \tag{8.24}$$

Now let γ be the geodesic of a freely-falling observer who enters the black hole outside of the region of collapsing matter. Choose the geodesic's affine parameter λ such that $\lambda = 0$ corresponds to the horizon crossing point $U = 0$. On this geodesic, $U(\lambda)$ is a smooth function of λ, so Taylor-expanding about $\lambda = 0$, we find that Eq. (8.24) reads

$$e^{-i\omega u} \approx e^{-i\omega\kappa^{-1}\ln(-\alpha\lambda)}, \tag{8.25}$$

where

$$\alpha = \left.\frac{dU}{d\lambda}\right|_{\lambda=0} \neq 0. \tag{8.26}$$

We see that a freely-falling observer indeed reports that the phase of $p_{\omega lm}$ has an oscillatory divergence at the event horizon.

We went to the trouble of illustrating this oscillatory divergence both because we will refer back to it in the following sections, but also in order to motivate using a geometric optics approximation to follow $p_{\omega lm}$ back from $\mathscr{I}^+$ to $\mathscr{I}^-$. Starting on $\mathscr{I}^+$ and propagating it backward, part of $p_{\omega lm}$ will be directly scattered off the gravitational field, and another part will enter the collapsing matter. Therefore, let $p_{\omega lm} = p^{(1)}_{\omega lm} + p^{(2)}_{\omega lm}$, where $p^{(1)}_{\omega lm}$ is the directly scattered part and $p^{(2)}_{\omega lm}$ is the part that enters the collapsing matter. It turns out that the $p^{(1)}$ part only contributes a term proportional to $\delta(\omega - \omega')$ in $\alpha_{\omega\omega'}$ and $\beta_{\omega\omega'}$, which does not result in any particle production, so, let us focus on the $p^{(2)}$ part.

This is where the geometric optics approximation is useful. On $\mathscr{I}^+$, $p^{(2)}$ has an oscillatory profile that oscillates more and more quickly as $u \to \infty$, i.e., near the event horizon. Since the oscillation frequency remains large near the event horizon throughout spacetime, the surfaces of constant phase of $p^{(2)}$ will be null surfaces near the event horizon. In other words, for large values of u, we can pull the waveform of $p^{(2)}$ on $\mathscr{I}^+$ back through the collapsing matter to a waveform on $\mathscr{I}^-$ by connecting corresponding points with null surfaces, as illustrated in Fig. 8.2 below. This is the geometric optics approximation. As long as the oscillation frequency is large compared to the local curvature, surfaces of constant phase are (very nearly) null surfaces, with the approximation getting better and better the closer one is to the event horizon.

Let ρ be a null generator of the event horizon. Continue this generator back through the coordinate origin (where the event horizon ends) toward the past so that it intersects $\mathscr{I}^-$. For convenience choose the null time v on $\mathscr{I}^-$ so that this intersection point happens at $v = 0$. Values of $p^{(2)}$ for points $u(v)$ near the event horizon on $\mathscr{I}^+$ will therefore be mapped to points $v < 0$ for small enough negative v. For $v > 0$,

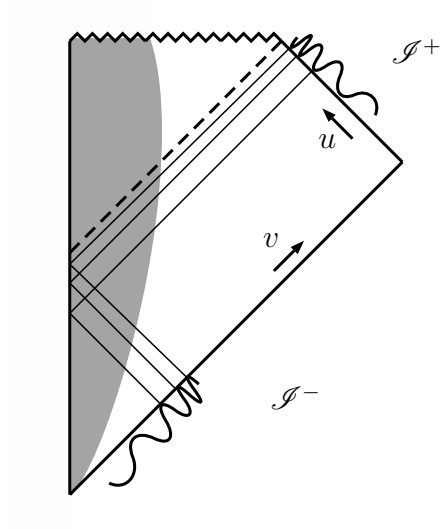

Figure 8.2: Illustration of the geometric optics approximation. This figure illustrates how troughs in a waveform at $\mathscr{I}^+$ are mapped onto corresponding troughs of a waveform at $\mathscr{I}^-$ via null rays.

the waveform will be vanishing.[4] Therefore, on $\mathscr{I}^-$ near $v = 0$, we conclude that $p^{(2)}$ is approximately given by

$$p^{(2)}_{\omega lm} \approx \begin{cases} \dfrac{1}{\sqrt{2\pi\omega}} \dfrac{1}{r} P^-_{\omega l} e^{-i\omega u(v)} Y^m_l(\theta, \phi) & v < 0 \\ 0 & v > 0 \end{cases} . \tag{8.27}$$

In the above, $P^-_{\omega l} \equiv P_{\omega l}(2M)$, i.e., the value of $P_{\omega l}(r)$ on the event horizon.

What is the correspondence between points $v < 0$ on $\mathscr{I}^-$ and points $u(v)$ on $\mathscr{I}^+$? These points are connected by null geodesics σ that remain close to the horizon. Therefore, we can characterize them using the geodesic deviation vector η^a that points from ρ to σ (Fig. 8.3). In particular, since $-\eta^a$ coincides with generators of v translations on $\mathscr{I}^-$, the functional dependence of $u(v)$ on v will be the same as how u depends on the affine parameter of a geodesic that is generated by $-\eta^a$. However, we already worked this out in Eq. (8.25) above! Therefore, it follows that on $\mathscr{I}^-$

[4] Alternatively, if we instead think about sending a waveform from $\mathscr{I}^-$ toward the future, points with $v > 0$ fall into the black hole and do not make it to $\mathscr{I}^+$.

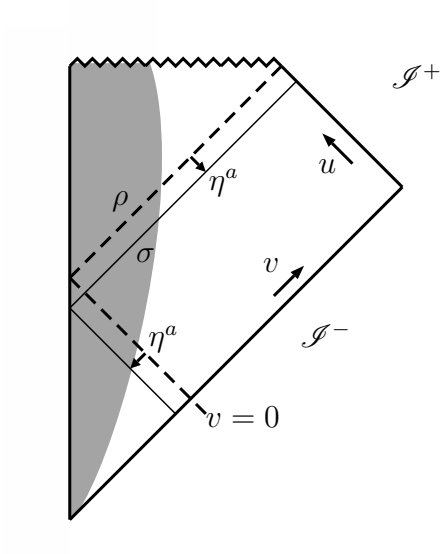

Figure 8.3: Geodesic deviation from a horizon generator, ρ, to a null geodesic, σ. The horizon generator is continued through $r = 0$ and out to $\mathscr{I}^-$. The null time where ρ hits $\mathscr{I}^-$ is chosen to be $v = 0$.

near $v = 0$,

$$p^{(2)}_{\omega lm} \approx \begin{cases} \dfrac{1}{\sqrt{2\pi\omega}} \dfrac{1}{r} P^-_{\omega l} e^{i\omega\kappa^{-1}\ln(-\alpha v)} Y^m_l(\theta,\phi) & v < 0 \\ \\ 0 & v > 0 \end{cases} . \tag{8.28}$$

We can now deduce the coefficients $\alpha^{(2)}_{\omega\omega'}$ and $\beta^{(2)}_{\omega\omega'}$ by Fourier transforming this expression for $p^{(2)}_{\omega lm}$ with respect to v. (Note that here we have also defined $\alpha_{\omega\omega'} = \alpha^{(1)}_{\omega\omega'} + \alpha^{(2)}_{\omega\omega'}$, and similarly for $\beta_{\omega\omega'}$.) According to the expansion (8.20), on $\mathscr{I}^-$ we have that

$$\begin{aligned} p^{(2)}_{\omega lm} = \int_0^\infty d\omega' \Bigg[& \alpha^{(2)}_{\omega\omega'} \left(\frac{1}{\sqrt{2\pi\omega'}} \frac{1}{r} \underbrace{F_{\omega' l}(r)}_{\to 1} e^{-i\omega' v} Y_{lm} \right) \\ & + \beta^{(2)}_{\omega\omega'} \left(\frac{1}{\sqrt{2\pi\omega'}} \frac{1}{r} \underbrace{F_{\omega' l}(r)}_{\to 1} e^{i\omega' v} \bar{Y}_{l(-m)} \right) \Bigg] \end{aligned} \tag{8.29}$$

and so it follows that

$$\alpha^{(2)}_{\omega\omega''} = \frac{1}{2\pi} \left(\frac{\omega''}{\omega} \right)^{1/2} P^-_{\omega l} \int_0^\infty dv \; e^{i\omega\kappa^{-1}\ln(-\alpha v) + i\omega'' v} \tag{8.30}$$

$$\beta^{(2)}_{\omega\omega''} = \frac{1}{2\pi} \left(\frac{\omega''}{\omega} \right)^{1/2} P^-_{\omega l} \int_0^\infty dv \; e^{i\omega\kappa^{-1}\ln(-\alpha v) - i\omega'' v} . \tag{8.31}$$

These integrals may evaluated by any handful of methods (contour integration, convolving multiple Fourier transforms, ...), and an explicit expression for them may be found in Ref. [43] for example. However, it turns out that $|\beta^{(2)}_{\omega\omega''}|^2 \propto (\omega'')^{-1}$, and so Eq. (8.22) is divergent. One way to overcome this divergence is to quantize in terms of wave packets. Nevertheless, we can immediately deduce the thermal form of $\langle 0^-|b^\dagger_{\omega lm} b_{\omega lm}|0^-\rangle$ at the expense of a bit of mathematical rigour with the following slick trick.

In particular, all we need is the relation [43]

$$|\alpha^{(2)}_{\omega\omega'}| = e^{\pi\omega\kappa^{-1}}|\beta^{(2)}_{\omega\omega'}| \tag{8.32}$$

as well as the distributional identity

$$\delta(\omega-\omega') = \int d\omega'' \left(\bar{\alpha}_{\omega\omega''}\alpha_{\omega'\omega''} - \bar{\beta}_{\omega\omega''}\beta_{\omega'\omega''}\right). \tag{8.33}$$

This last identity may be established by inserting Eq. (8.21) into the commutation relation $[b^\dagger_\omega, b^\dagger_{\omega'}] = \delta(\omega-\omega')$. Then, we formally have that

$$\begin{aligned}\lim_{\omega'\to\omega}\delta(\omega-\omega') &\stackrel{!}{=} \int d\omega'' \left(|\alpha_{\omega\omega''}|^2 - |\beta_{\omega\omega''}|^2\right) \\ &\stackrel{!}{=} \int d\omega'' (e^{2\pi\omega\kappa^{-1}}-1)|\beta_{\omega\omega''}|^2\,,\end{aligned}$$

whence, brutally,

$$\int_0^\infty d\omega'\ |\beta_{\omega\omega'}|^2 \stackrel{!}{=} \frac{\delta(0)}{e^{2\pi\omega\kappa^{-1}}-1}\,. \tag{8.34}$$

Therefore, barring the unpalatable divergence, we already see that the spectrum of mode occupation at $\mathscr{I}^+$ is Planckian with a temperature $\kappa/2\pi$.

Alternatively, one may arrive at the Planckian spectrum of radiation more rigorously by arguing using wave packets. Hawking discuss this approach in detail, but here we will at least describe the prescription for building and working with wave packets. Following Ref. [43] (with the appropriate modifications due to the opposite convention for positive frequency modes), define

$$p^{(2)}_{jn} = \frac{1}{\sqrt{\epsilon}}\int_{j\epsilon}^{(j+1)\epsilon} d\omega\ e^{i\omega(2\pi n/\epsilon)}p^{(2)}_\omega, \tag{8.35}$$

where we have suppressed the lm subscripts to avoid clutter. Here, $j, n \in \mathbb{Z}$, $j \geq 0$, and $\epsilon > 0$. Defined as such, the $p^{(2)}_{jn}$ are wave packets of frequency $\sim jn$ and width $\sim \epsilon^{-1}$ that are centered on the null time $u = 2\pi n/\epsilon$. For illustration, $|p^{(2)}_{jn}|$ is plotted as a function of u in Fig. 8.4 for $p^{(2)}_{\omega}$ set equal to $e^{-i\omega u}$. From their definition, it follows that

$$p^{(2)}_{jn} = \int_0^\infty d\omega' \left(\alpha^{(2)}_{jn\omega'} f_{\omega'} + \beta^{(2)}_{jn\omega'} \bar{f}_{\omega'} \right) , \tag{8.36}$$

where

$$\alpha^{(2)}_{jn\omega'} = \frac{1}{\sqrt{\epsilon}} \int_{j\epsilon}^{(j+1)\epsilon} d\omega \; e^{i\omega(2\pi n/\epsilon)} \alpha^{(2)}_{\omega\omega'} \tag{8.37}$$

and similarly for $\beta^{(2)}_{jn\omega'}$. The field is then correspondingly quantized as

$$\varphi = \sum_{j=0}^{\infty} \sum_{n=-\infty}^{\infty} p_{jn} b_{jn} + \bar{p}_{jn} b^\dagger_{jn} \; + \; (q \text{ terms}). \tag{8.38}$$

(We have still suppressed the lm indices above.) Note that the p_{jnlm} defined as such are still orthonormal with respect to (8.12), which can be easily checked by calculating $(p_{jnlm}, p_{j'n'l'm'})$ and using the (continuum) orthonormality of the $p_{\omega lm}$.

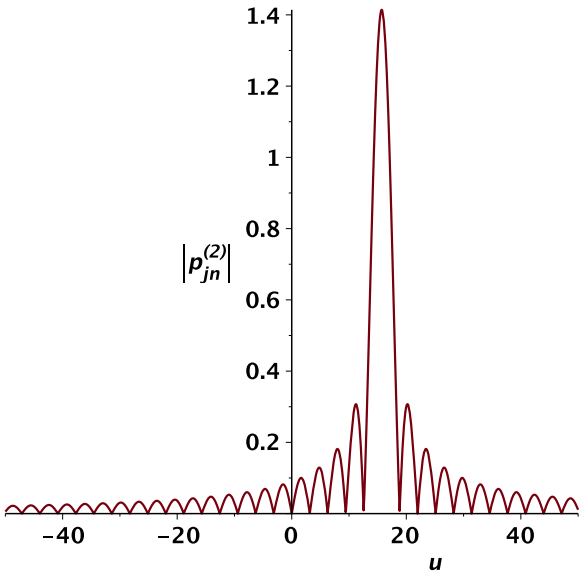

Figure 8.4: A plot of $|p^{(2)}_{jn}|$ for $p^{(2)}_{\omega}$ set equal to $e^{-i\omega u}$. Plot parameters are $\epsilon = 2$, $j = 4$, and $n = 5$.

At this point, $\alpha^{(2)}_{jn\omega'}$ and $\beta^{(2)}_{jn\omega'}$ still obey the important property [43]

$$|\alpha^{(2)}_{jn\omega'}| = e^{\pi\omega\kappa^{-1}} |\beta^{(2)}_{jn\omega'}|, \tag{8.39}$$

which ultimately lets one compute

$$\langle 0^-|b^\dagger_{jn}b_{jn}|0^-\rangle = \int_0^\infty d\omega' \ |\beta_{jn\omega'}|^2, \tag{8.40}$$

once more yielding a thermal distribution. The calculation is very similar to the mathematically perverse trick discussed above, so we refer the reader to Ref. [43] for further details.

We end this section by reminding ourselves that the calculations above all stemmed from approximating the propagation of the mode functions back through the bulk spacetime from $\mathscr{I}^+$ to $\mathscr{I}^-$. The geometric optical approximation is increasingly accurate near the event horizon, which, in the wave packet formalism, means that the approximate calculation above is increasingly exact for wavepackets that are centered about late null times u on $\mathscr{I}^+$. A striking implication of this observation is that the spectrum of late radiation is insensitive to the details of the collapse—the spectrum is almost exactly thermal. The full spectrum of Hawking radiation is not thermal, however, instead receiving ω- and l-dependent corrections in the form of greybody factors. Further discussion of greybody factors may be found in [45] and [230, 231].

8.2 Eternal black holes

In the derivation of Hawking radiation discussed above, the collapsing matter was absolutely crucial for finding particle creation in the asymptotic future. However, we saw that the late-time behaviour of the radiation was insensitive to the precise details of the collapse, which suggests (as Birrell and Davies also note [192]) that the presence of radiation is mostly due to the spacetime topology and the presence of a horizon. Therefore, while on one hand it may even seem antithetical to look for Hawking radiation on an *eternal* black hole background, we will see that we can deduce the thermal properties of the radiation via judicious choices of the vacuum state.[5]

[5]This section only paints an overview of the most important points for setting up the black hole information problem. We will not go into the same level of mathematical detail as the previous section. Instead, Refs. [192, 229] are good resources for further calculational details.

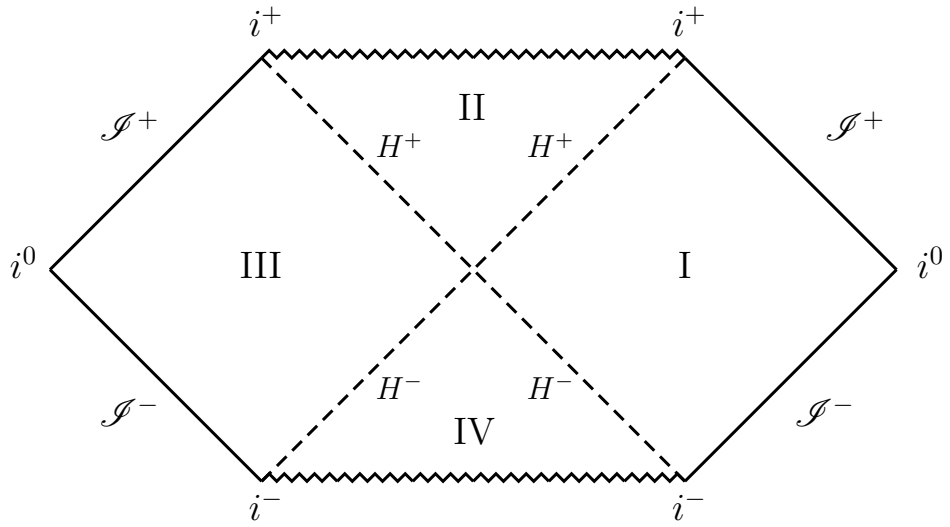

Figure 8.5: Penrose diagram for the maximally extended Schwarzschild spacetime. Regions I and III are asymptotically flat regions outside the black hole (region II) and the white hole (region IV). The black hole horizon is labelled H^+ and the white hole horizon is labelled H^-.

Here we will consider a scalar field on a maximally-extended Schwarzschild background, for which the relevant Penrose diagram is shown in Fig. 8.5.

Recall that Kruskal-Szekeres coordinates constitute a coordinate system that covers the entire maximally-extended manifold. These are coordinates (U, V, θ, ϕ) given by

$$U = -e^{-u/4M} \tag{8.41}$$

$$V = e^{v/4M} \tag{8.42}$$

(where, as before, $u = t - r_*$ and $v = t + r_*$) in which the metric reads

$$ds^2 = -32M^3 \frac{e^{-r/2M}}{r} dU dV. \tag{8.43}$$

(A good resource for further discussion of this coordinate system is Ch. 6.4 of Ref. [160].) Similarly to Eqs. (8.4) and (8.5), if one lets $T = \frac{1}{2}(U+V)$, $X = \frac{1}{2}(V-U)$, and then takes as an ansatz

$$\varphi(T, X, \theta, \phi) = \frac{1}{r} h(T, X) Y_l^m(\theta, \phi), \tag{8.44}$$

the resulting differential equation for $h(T, X)$ is

$$-\frac{\partial^2 h}{\partial T^2} + \frac{\partial^2 h}{\partial X^2} - \frac{1}{r}\left[32M^3 e^{-r/2M}\left(\frac{l(l+1)}{r^2} + \mu^2\right) - \frac{\partial^2 r}{\partial T^2} + \frac{\partial^2 r}{\partial X^2}\right] h = 0\,. \tag{8.45}$$

One can therefore define a set of mode functions $f^K_{\omega lm} \sim e^{-i\omega U}, e^{-i\omega V}$ that are regular everywhere on the spacetime and use these to quantize the scalar field:

$$\varphi = \sum_{lm} \int d\omega \left(f^K_{\omega lm} a^K_{\omega lm} + \bar{f}^K_{\omega lm} (a^K_{\omega lm})^\dagger \right) . \tag{8.46}$$

This choice of mode functions defines the Kruskal vacuum (also known as the Hartle-Hawking state) via the condition

$$a^K_{\omega lm} |0_K\rangle = 0 \quad \forall\ \omega, l, m. \tag{8.47}$$

We could also choose to expand the field using mode functions whose positive frequency part behaves like $e^{-i\omega u}$ and $e^{-i\omega v}$ at $\mathscr{I}^+$ and $\mathscr{I}^-$ as before. Again, the price we pay is that these mode functions are singular on both the future black hole horizon (H^+) and past white hole horizon (H^-); however, these are the modes that describe free spherical waves at $\mathscr{I}^-$ and $\mathscr{I}^+$. Note that since the maximally extended Schwarzschild spacetime has two copies each of $\mathscr{I}^-$ and $\mathscr{I}^+$, we should double the set of modes, meaning that the quantized field is given by

$$\varphi = \sum_{lm} \int d\omega \left(p^{\mathrm{I}}_{\omega lm} b^{\mathrm{I}}_{\omega lm} + \bar{p}^{\mathrm{I}}_{\omega lm} (b^{\mathrm{I}}_{\omega lm})^\dagger + p^{\mathrm{III}}_{\omega lm} b^{\mathrm{III}}_{\omega lm} + \bar{p}^{\mathrm{III}}_{\omega lm} (b^{\mathrm{III}}_{\omega lm})^\dagger \right) . \tag{8.48}$$

As such, $b^{\mathrm{I}}_{\omega lm}$ and $(b^{\mathrm{I}}_{\omega lm})^\dagger$ annihilate and create excitations in the asymptotically flat region I, while the other creation and annihilation operators act in the region III. The state that is annihilated by all of the $b^{\mathrm{I}}_{\omega lm}$ and $b^{\mathrm{III}}_{\omega lm}$ is known as both the Schwarzschild vacuum and the Boulware vacuum, which here we denote by $|0_S\rangle$.

The strategy is then very similar to that for the collapsing black hole. One again writes

$$p^{\mathrm{I}}_{\omega lm} = \int d\omega \left(\alpha^{\mathrm{I}}_{\omega\omega'} f^K_{\omega lm} + \beta_{\omega\omega'} \bar{f}^K_{\omega lm} \right) \tag{8.49}$$

and once more arrives at the conclusion that $\langle 0_K | (b^{\mathrm{I}}_{\omega lm})^\dagger b^{\mathrm{I}}_{\omega lm} | 0_K \rangle$ has a thermal distribution. One could argue that the Kruskal vacuum $|0_K\rangle$ is the "natural" vacuum for the maximally extended spacetime simply on the grounds that its modes are regular everywhere except the singularities. However, one can also show that this vacuum also results in an ingoing flux of thermal radiation at $\mathscr{I}^-$. The Kruskal vacuum

therefore describes a situation in which the eternal black hole is in equilibrium with a bath of thermal radiation, which is perhaps a more compelling argument that $|0_K\rangle$ should be the "natural" vacuum.

Finally, one can show that it is possible to write the Kruskal vacuum in terms of the Fock basis states generated by acting with $(b^{\mathrm{I}}_{\omega lm})^\dagger$ and $(b^{\mathrm{III}}_{\omega lm})^\dagger$ on the Schwarzschild vacuum [192]:

$$|0_K\rangle = \bigotimes_{\omega lm} \left[\sqrt{1 - e^{-2\pi\omega\kappa^{-1}}} \sum_{n=0}^{\infty} |n\rangle^{\mathrm{I}}_{\omega lm} |n\rangle^{\mathrm{III}}_{\omega l(-m)} \right] . \tag{8.50}$$

In the equation above, $|n\rangle^{\mathrm{I}}_{\omega lm}$ (resp. $|n\rangle^{\mathrm{III}}_{\omega lm}$) denotes the state obtained by acting n times with the creation operator $(b^{\mathrm{I}}_{\omega lm})^\dagger$ (resp. $(b^{\mathrm{III}}_{\omega lm})^\dagger$) on $|0_S\rangle$. In particular, note that the state of each mode ωlm is entangled across the regions I and III.

The last observation that we will make is that we could equally well repeat the same analysis with wave packets like those discussed in the last section. Explicitly, we can define

$$p^{\mathrm{I}}_{jn} = \frac{1}{\sqrt{\epsilon}} \int_{j\epsilon}^{(j+1)\epsilon} d\omega \; e^{i\omega(2\pi n/\epsilon)} p^{\mathrm{I}}_{\omega}, \tag{8.51}$$

with p^{III}_{jn} defined similarly. It immediately follows that we can write $|0_K\rangle$ in terms of these wave packets instead so that

$$|0_K\rangle \sim \bigotimes_{jnlm} \sum_{\nu=0}^{\infty} e^{-\pi(j\epsilon)\nu\kappa^{-1}} |\nu\rangle^{\mathrm{I}}_{jnlm} |\nu\rangle^{\mathrm{III}}_{jnl(-m)} \tag{8.52}$$

and where the p^{I}_{jn} are localized near the horizon for large n.

8.3 The Black Hole Information Problem

So far we have only discussed the dynamics of a quantum field on a black hole spacetime background without taking into account the backreaction of the field on the geometry. One can show that the Hawking radiation results in a positive energy flux out to $\mathscr{I}^+$ that is balanced by a negative energy flux into the black hole [192]. Therefore, if we take backreaction into account, we conclude that the emission of Hawking radiation will cause the black hole to slowly lose its mass over time and eventually disappear. While such a process is decidedly dynamical, we should still

be able to trust the picture built up from studying field theory on a fixed black hole background because the evaporation process is so slow. In other words, to a very good approximation, one should be able to think of a dynamical evaporating black hole as being instantaneously stationary (at least until the black hole's mass is on the order of the Planck scale) and be able to extrapolate based on the calculations of the previous two sections.

An unfortunate consequence of this picture is that the complete evaporation of a black hole is at odds with the unitarity of quantum mechanics. Consider again the black hole formed from the collapse of matter. Suppose that the collapsing matter is massive and starts out in some pure state at i^- and that we have the (pure) ingoing vacuum on $\mathscr{I}^-$. If the black hole fully evaporates so that the spacetime approaches a Minkowski configuration in the future (or in other words, so that there is no event horizon, as shown in Fig. 11.1), then our calculation of Hawking radiation suggests that the final state on $\mathscr{I}^+$ (and also on i^+ if we have massive fields) is a *mixed* state that is nearly thermal. The process of black hole formation and evaporation therefore maps pure states into mixed states, and so we conclude that black holes can destroy coherent quantum information. This is the original black hole information problem.

The "original resolution" of the black hole information problem, as was suggested by Page [228] among others, is to assert that the final state of the Hawking radiation is actually pure, and that it only appears approximately thermal (e.g., with respect to any local probe). After all, a sequence of instantaneously stationary black holes can only be an approximation to the full dynamical black hole. In this picture, quantum information is preserved via subtle correlations in the state of the radiation field.

This idea was formalized by the principle of black hole complementarity [23]. According to complementarity, unitarity is preserved by maintaining that, from the point of view of an observer who remains outside the black hole, no quantum information actually crosses the apparent horizon. Instead, it appears encoded in a structure called the "stretched horizon" [23, 133] that is located just above the black hole's

apparent horizon and whose boundary in spacetime is timelike. The stretched horizon has complicated dynamics, but is responsible for gradually returning quantum information to outgoing Hawking radiation in this picture.

At the same time, an observer who tries to enter the black hole sees no stretched horizon, nor anything that they would deem "out of the ordinary" based on their knowledge of general relativity and quantum field theory on a curved spacetime background. In particular, they would report than any system that fell into the black hole, as well as the quantum information that it carries, is still there behind the horizon, provided that the observer and the system in question are still far away from the singularity. To the infalling observer, the black hole interior is very much a real place that can store quantum information. In other words, the stationary observer and the infalling observer have *complementary* descriptions of where and how quantum information is stored in spacetime. There are no information theoretic contradictions (such as violations of no-cloning) because an infalling observer who crosses the apparent horizon cannot communicate the results of any measurements they make to an exterior observer.

The modern black hole information problem consists of an inconsistency with the picture of complementarity described above that was identified by Almheri, Marolf, Polchinski, and Sully (collectively known as "AMPS") [18, 19]. More precisely, they argued that the following four postulates cannot all be true at the same time:

1. *Unitarity* — As viewed by an observer who remains far away from the black hole, the formation and evaporation of the hole is a unitary quantum-mechanical process.

2. *Local Effective Field Theory* — To the exterior of the black hole's stretched horizon, the physics of matter is well described by a local effective field theory on a black hole spacetime background.

3. $S_{\rm bh} = S_{\rm BH}$ — As viewed by an observer who remains far away from the horizon, the black hole is a quantum-mechanical system that is represented by a finite

dimensional Hilbert space. Moreover, the von Neumann entropy of an old black hole, $S_{\rm bh}$, is (if not exactly, approximately) equal to its Bekenstein-Hawking entropy, $S_{\rm BH}$.

4. *No Drama* — An observer who crosses the apparent horizon of the black hole (but remains far from its central singularity) encounters nothing that runs contrary to the predictions of semiclassical general relativity and effective field theory.

In my view, the most robust way to argue that these four postulates are inconsistent is to construct a violation of strong subadditivity of von Neumann entropy (SSA). The following construction makes use of the wave packets discussed earlier and constitutes an attempt on my part to be as careful as possible in defining the relevant Hilbert space factors.

Consider a quantization of a massless scalar field on an evaporating black hole background in terms of wave packets of the form given in Eq. (8.35). On $\mathscr{I}^+$, pick a particular wave packet jn_0lm, where n_0 is chosen so that the wave packet is excited late during the black hole's lifetime. Denote the Fock space of this mode by B. Similarly, let R denote the Hilbert space factor corresponding to the joint Fock space of all those modes for which $n < n_0$. These spaces B and R are careful definitions of the "late" and "early" radiation that are referenced in the literature.

On $\mathscr{I}^+$, the mode B is localized about $u = 2\pi n_0/\epsilon$, and it remains localized and near the horizon when propagated back toward the black hole. According to postulates 2 and 4, that an infalling observer sees nothing out of the ordinary as they cross the horizon means that the state of the field should be in a state that is extremely close to the Kruskal vacuum, in analogy with the eternal black hole analysis. In particular, the crucial requirement is that B must be entangled with some other mode A on the other side of the horizon. Moreover, based on the eternal black hole case, we should expect the joint state of A and B to be a pure state. In the Kruskal case, this joint

state would be, up to relative weights in individual terms,

$$|\psi\rangle_{AB} \sim \sum_{\nu} |\nu\rangle^{\mathrm{I}}_{jn_0lm} |\nu\rangle^{\mathrm{III}}_{jn_0l(-m)} \,. \tag{8.53}$$

Now let us assess the entropies that go into the strong subadditivity (SSA) relation for these three systems:

$$S_{AB} + S_{BR} \geq S_B + S_{ABR} \,. \tag{8.54}$$

If AB is pure, then it follows that $S_{AB} = 0$ and $S_{ABR} = S_R$. According to postulate 1, if the final state resulting from black hole evaporation is pure and B is chosen to be late enough, it must be that B purifies the early radiation R. Or, in terms of entropies, it must be that $S_{BR} < S_R$. Putting these conditions together, however, leads to a contradiction:

$$\begin{aligned} S_B + S_{ABR} &= S_B + S_R && (2) + (4) \\ &\leq S_{AB} + S_{BR} && (\mathrm{SSA}) \\ &= 0 + S_{BR} && (2) + (4) \\ &< S_R && (1) \,. \end{aligned} \tag{8.55}$$

It cannot be that $S_B + S_R < S_R$, so we have an inconsistency.

Despite being as careful as possible, there are still plenty of objections that can be made about the construction above. For example, most of the entropies above are divergent, and so a better treatment would be more careful with regulating these divergences. Or, the conclusions drawn about the entropies of A and B were based on facts about the Kruskal vacuum, so one should really quantify to what extent $S_{AB} \approx 0$ and $S_{ABR} \approx S_R$ are good approximations.

A different sort of objection is that Schwarzschild modes of the form $p_{\omega lm}$ or wave packets p_{jn} are physically unreasonable from the perspective of any infalling observer. Naïvely, Schwarzschild modes have an oscillatory divergence at the horizon, but even if one deals with this divergence by constructing a wave packet that vanishes at the horizon, any wave packet with a finite width centered on a fixed frequency asymptotically will receive an enormous blueshift near the horizon. Such a blueshift

seems unreasonable from the perspective of an effective field theory that an infalling observer might try to write down. Nevertheless, that no drama seems to demand an entanglement structure like that of Eq. (8.52) in a basis that is appropriate for asymptotic observers is still problematic as an "in principle" statement.

Be that as it may, the prevailing opinion in the community is that in order to resolve the modern black hole information problem, one must either abandon or invalidate one or more of the postulates (1)-(4) above. "Firewalls" refers to a scenario in which one abandons the 4^{th} postulate. Namely, one can preserve unitarity by letting A and B be unentangled, but this results in a large expectation value for the local stress-energy tensor at the horizon [45]. The resulting interpretation is that a energetic curtain of radiation—the firewall—lurks just behind the event horizon, violating postulate (4) in a flagrant way.

While Chapter 9 is concerned with the task of recovering a single qubit from a black hole, Chapters 10 and 11 examine two ways in which AMPS' postulates may be relaxed or rebutted.

Chapter 9

HOW TO RECOVER A QUBIT THAT HAS FALLEN INTO A BLACK HOLE

We demonstrate an algorithm for the retrieval of a qubit, encoded in spin angular momentum, that has been dropped into a no-firewall black hole. Retrieval is achieved analogously to quantum teleportation by collecting Hawking radiation and performing measurements on the black hole. Importantly, these methods only require the ability to perform measurements from outside the event horizon.

This chapter was published as Ref. [7], A. Chatwin-Davies, A. S. Jermyn, and S. M. Carroll, "How to recover a qubit that has fallen into a black hole," Phys. Rev. Lett. ***115*** *(2015) 261302,* `arXiv:1507.03592`.

9.1 Introduction

Recovering the complete quantum state of a black hole from the Hawking radiation [43] into which it evaporates is notoriously difficult [45]. In this letter we tackle a simpler problem: recovering the quantum state of a single spin qubit that has fallen into an evaporating black hole.

Our protocol uses information about the spin state of the black hole before and after the qubit entered, as well as the state of pairs of Hawking particles. The outline of the procedure, sketched in Fig. 9.1, is as follows:

1. The initial spin state of the black hole is measured, putting the density matrix of the black hole in the form $\rho_B = \rho_B^{(\text{int})} \otimes |j,m\rangle\langle j,m|$, where j, m are the quantum numbers for total and projected angular momentum, and $\rho_B^{(\text{int})}$ characterizes the internal degrees of freedom. Perfect fidelity can be achieved only if $m = 0$; the experimenter can measure the spin along different axes until this outcome is attained.

2. The experimenter collects a single Hawking photon that is part of a Bell pair, the other photon of which falls into the hole.

3. The qubit, a photon in an arbitrary helicity state $|\phi\rangle_A = \alpha|\epsilon^+\rangle_A + \beta|\epsilon^-\rangle_A$, is dropped into the hole.

4. The black hole's spin state is again measured, so that the density matrix becomes $\rho'_B = \rho'^{(\mathrm{int})}_B \otimes |j', m'\rangle\langle j', m'|$. Dephasing of the hole's spin does not occur if the interactions between the hole's spin and its internal state are rotationally-invariant (conserve angular momentum).[1]

5. The initial state of the qubit can then be reconstructed from the state of the collected Hawking photon.

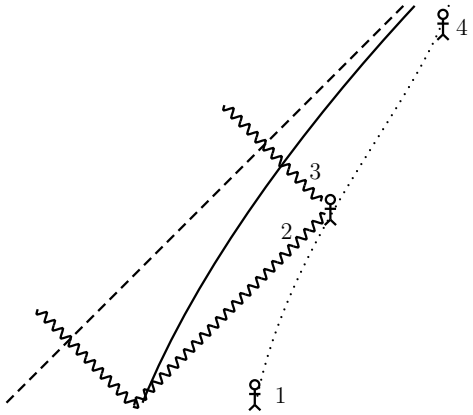

Figure 9.1: Sketch of the qubit recovery protocol on a Penrose diagram. The numbers correspond to the steps enumerated above (details of the initial measurement 1. are not shown). The dashed line represents the event horizon, the solid line represents the stretched horizon [23], and the dotted line represents the experimenter's trajectory.

This falls far short of a resolution to the information-loss problem [40,233–235], but it does provide a concrete illustration of how information can escape from a black hole in certain special circumstances, and is similar in spirit to earlier discussions about using conserved quantities to recover black hole information [236, 237]. Moreover, whether or not the Page time [41] has elapsed does not affect information recovery,

[1]Concretely, suppose that there was some conditional interaction between the black hole's internal degrees of freedom and its spin which would take a state $|BH\rangle \otimes (\alpha|\epsilon^+\rangle + \beta|\epsilon^-\rangle)$ to a state $\alpha|BH^+\rangle \otimes |\epsilon^+\rangle + \beta|BH^-\rangle \otimes |\epsilon^-\rangle$, where $\langle BH^+|BH^-\rangle = 0$. If, for example, $\alpha = \beta = 1/\sqrt{2}$, then angular momentum in the x direction would not be conserved by the interaction.

since the protocol is not concerned with reconstructing the state of the black hole. In this regard the protocol is entirely distinct from the Hayden-Preskill result [238].

9.2 A protocol for retrieving individual qubits

Suppose that Alice sits outside a black hole and has in her possession a photon in some state $|\phi\rangle_A = \alpha|\epsilon^+\rangle_A + \beta|\epsilon^-\rangle_A$ that is unknown to her. Here, the basis states $|\epsilon^+\rangle_A$ and $|\epsilon^-\rangle_A$ represent the photon's helicity, and thus have angular momentum projection $+1$ and -1 respectively. First, Alice measures the black hole's angular momentum and finds it in the state $|j, m\rangle_B$. (We suppress the state of the black hole's internal degrees of freedom, which will play no role in our analysis.) Such a measurement is technologically formidable, but one which Alice could in principle perform with the help of a sufficiently large Stern-Gerlach apparatus or by carefully measuring frame dragging.

Before dropping her qubit into the black hole, Alice collects a single Hawking photon. We assume that the emitted photon is one half of a pair, the other one of which falls into the hole. We also assume that Alice completes the protocol before any more Hawking particles are emitted. The pairs of particles will have equal mass and opposite gauge and Poincaré quantum numbers. Let us focus on angular momentum.

The states of photons with definite angular momentum are spherical waves that may be labelled by the quantum numbers for linear momentum, $k \in (0, \infty)$; total (spin plus orbital) angular momentum, $\eta \in \{1, 2, \dots\}$; projected angular momentum, $\mu \in \{-\eta, ..., \eta\}$; and parity, $\bar{\omega} \in \{+1, -1\}$ [239]. We assume that the photons are each produced in the lowest angular momentum state ($\eta = 1$) since this is the dominant mode of Hawking photon production. Alternatively, Alice can measure her photon's total angular momentum and then discard her photon and restart the protocol if it does not have $\eta = 1$. In order to preserve CPT, the two photons are produced with the same parity, since they are uncharged and since the wavefunctions of different parity for each $(k\eta\mu)$ have the same sign under T. The photons must also be created in a zero total angular momentum state to conserve angular momentum. As such,

after Alice measures the parity of her photon, the angular momentum of the ingoing (i) and outgoing (o) Hawking photons is

$$|0,0\rangle_{io} \equiv \frac{1}{\sqrt{3}}\left(|1,1\rangle_i|1,-1\rangle_o + |1,-1\rangle_i|1,1\rangle_o - |1,0\rangle_i|1,0\rangle_o\right). \tag{9.1}$$

(Further justification for this model is provided in the next section.)

Next, Alice measures the squared projected angular momentum of her photon. If she obtains the result $\mu^2 = 0$, then she discards her photon and restarts the protocol. Otherwise, the ingoing and outgoing photons are projected into the Bell state $|\Phi\rangle_{io} = (|1,1\rangle_i|1,-1\rangle_o + |1,-1\rangle_i|1,1\rangle_0)/\sqrt{2}$. Finally, Alice drops in her qubit, and then measures the angular momentum of the hole again, determining it to be $|j',m'\rangle_B$.

After Alice collects a suitable Hawking photon and drops her qubit into the black hole, the total state of the black hole and the three photons is therefore $|\Psi\rangle = |j,m\rangle_B \otimes |\phi\rangle_A \otimes |\Phi\rangle_{io}$. Alice is ignorant of what happens inside the black hole. What Alice can know, however, is the total angular momentum of the black hole and the projection of its angular momentum vector along some axis. As such, let us rewrite the AiB subsystem in the total angular momentum basis:

$$\begin{aligned}
|\Psi\rangle = \frac{1}{\sqrt{2}}\Bigg\{ &\sum_{\sigma=-2}^{2}\Big[\langle {}^{j}_{m}\,{}^{2}_{2}|^{j+\sigma}_{m+2}\rangle |j+\sigma, m+2\rangle \otimes \alpha|1,-1\rangle_o \\
&\qquad + \langle {}^{j}_{m}\,{}^{2}_{-2}|^{j+\sigma}_{m-2}\rangle |j+\sigma, m-2\rangle \otimes \beta|1,1\rangle_o\Big] \\
&+ \frac{1}{\sqrt{6}}\sum_{\sigma=-2}^{2}\left[\langle {}^{j}_{m}\,{}^{2}_{0}|^{j+\sigma}_{m}\rangle |j+\sigma,m\rangle \otimes (\alpha|1,1\rangle_o + \beta|1,-1\rangle_o)\right] \\
&+ \frac{1}{\sqrt{2}}\sum_{\delta=-1}^{1}\left[\langle {}^{j}_{m}\,{}^{1}_{0}|^{j+\delta}_{m}\rangle |j+\delta,m\rangle^{\perp} \otimes (\alpha|1,1\rangle_o - \beta|1,-1\rangle_o)\right] \\
&+ \frac{1}{\sqrt{3}}|j,m\rangle^{\vdash} \otimes (\alpha|1,1\rangle_o + \beta|1,-1\rangle_o)\Bigg\}.
\end{aligned} \tag{9.2}$$

The symbols $\langle {}^{j_1}_{m_1}\,{}^{j_2}_{m_2}|^{j}_{m}\rangle \equiv \langle j_1,m_1;j_2,m_2|j,m\rangle$ denote appropriate Clebsch-Gordan coefficients. We have also suppressed the label AiB on the total angular momentum kets. Note that some of the $\langle {}^{j_1}_{m_1}\,{}^{j_2}_{m_2}|^{j}_{m}\rangle$ could be zero. For now, we will assume that $-j+2 < m < j-2$. In particular note the following states: $|j,m\rangle$, which comes

from $j \otimes 2$; $|j,m\rangle^{\perp}$, which comes from $j \otimes 1$; and $|j,m\rangle^{\vdash}$, which comes from $j \otimes 0$. These states have the same angular quantum numbers, but are orthogonal.

Next, Alice queries the black hole's total angular momentum by performing the following orthogonal measurement on AiB:

$$\begin{aligned} \hat{F}_1 &= \sum_a |a,m\rangle\langle a,m|, \\ \hat{F}_2 &= \sum_a |a,m+2\rangle\langle a,m+2| + |a,m-2\rangle\langle a,m-2|, \\ \hat{F}_3 &= \hat{I}_{AiB} - \hat{F}_1 - \hat{F}_2. \end{aligned} \tag{9.3}$$

Note that by construction, only the results $\hat{F}_1$ and $\hat{F}_2$ may be obtained for black hole states which may emerge from this protocol. The protocol for retrieving the state $|\phi\rangle$ is then as follows:

Case 1: Alice obtains the result $\hat{F}_1$. In this case, the whole system collapses to a state that is proportional to the second and third lines of Eq. (9.2). Alice then measures the total angular momentum $\hat{J}^2$ of the black hole.

If Alice measures the result $J^2 = (j \pm 2)(j \pm 2 + 1)$, then she knows that the spin that she holds is in the desired state $|\phi\rangle_o = \alpha|1,1\rangle_o + \beta|1,-1\rangle_o$.

If Alice measures the result $J^2 = (j \pm 1)(j \pm 1 + 1)$, then the total system is in the state

$$\begin{aligned} |\Psi'\rangle \propto & \frac{1}{\sqrt{6}} \langle {}^{j}_{m}\, {}^{2}_{0} | {}^{j\pm 1}_{m} \rangle |j \pm 1, m\rangle \otimes |\phi\rangle_o \\ & + \frac{1}{\sqrt{2}} \langle {}^{j}_{m}\, {}^{1}_{0} | {}^{j\pm 1}_{m} \rangle |j \pm 1, m\rangle^{\perp} \otimes |\phi'\rangle_o, \end{aligned} \tag{9.4}$$

while if she measures the result $J^2 = j(j+1)$, then the total system is in the state

$$\begin{aligned} |\Psi'\rangle \propto & \left[\frac{1}{\sqrt{6}} \langle {}^{j}_{m}\, {}^{2}_{0} | {}^{j}_{m} \rangle |j,m\rangle + \frac{1}{\sqrt{3}} |j,m\rangle^{\vdash} \right] \otimes |\phi\rangle_o \\ & + \frac{1}{\sqrt{2}} \langle {}^{j}_{m}\, {}^{1}_{0} | {}^{j}_{m} \rangle |j,m\rangle^{\perp} \otimes |\phi'\rangle_o, \end{aligned} \tag{9.5}$$

where $|\phi'\rangle_o = \alpha|1,1\rangle_o - \beta|1,-1\rangle_o$. Each of these states represents a mixed density matrix for the spin that Alice holds unless some of the Clebsh-Gordan coefficients

vanish. In particular, some algebra reveals that

$$\begin{aligned} \langle {}^{j}_{m}\ {}^{2}_{0} | {}^{j+1}_{m} \rangle^2 &= \tfrac{3m^2(j+m+1)(j-m+1)}{j(j+1)(j+2)(2j+1)} \\ \langle {}^{j}_{m}\ {}^{2}_{0} | {}^{j-1}_{m} \rangle^2 &= \tfrac{3m^2(j+m)(j-m)}{j(j+1)(j-1)(2j+1)} \\ \langle {}^{j}_{m}\ {}^{1}_{0} | {}^{j}_{m} \rangle^2 &= \tfrac{m^2}{j(j+1)}. \end{aligned} \tag{9.6}$$

At the beginning of the protocol, Alice may measure j and determine if it is an integer. If not, she may repeatedly throw spin-1/2 particles into the black hole and measure j until she measures an integral value. She may then repeatedly measure the black hole's angular momentum projection along different axes until she obtains $m = 0$, before collecting a Hawking photon and tossing her qubit into the hole. In this way, the Clebsch-Gordan coefficients (9.6) may be made to vanish, allowing Alice to recover the qubit.

Case 2: Alice obtains the result $\hat{F}_2$. In this case, the whole system collapses to a state that is proportional to the first line of Eq. (9.2). Next, Alice measures the total angular momentum $\hat{J}^2$, obtaining the result $J^2 = (j+\sigma)(j+\sigma+1)$ for some $\sigma \in \{-2, \ldots, 2\}$. The total state is then

$$\begin{aligned} |\Psi''\rangle \propto\ & \alpha \, \langle {}^{j}_{m}\ {}^{2}_{2} | {}^{j+\sigma}_{m+2} \rangle |j+\sigma, m+2\rangle \otimes |1,-1\rangle_o \\ & - \beta \, \langle {}^{j}_{m}\ {}^{2}_{-2} | {}^{j+\sigma}_{m-2} \rangle |j+\sigma, m-2\rangle \otimes |1,1\rangle_o. \end{aligned} \tag{9.7}$$

We are faced with the problem of disentangling the AiB part of the system from the o part which Alice holds. She may accomplish this task with the help of a spin-2 ancilla and a local entangling unitary. Suppose Alice holds a spin-2 ancilla, A', that she prepares in the state $|2,0\rangle_{A'}$. If she then implements a local entangling unitary operator $U_{oA'}$ such that

$$\begin{aligned} U_{oA'}|1,1\rangle_o|2,0\rangle_{A'} &= |1,1\rangle_o|2,2\rangle_{A'} \\ U_{oA'}|1,-1\rangle_o|2,0\rangle_{A'} &= |1,-1\rangle_o|2,-2\rangle_{A'}, \end{aligned} \tag{9.8}$$

upon acting with $U_{oA'}$ on the spins that she holds, the total state

$$I_{AiB} \otimes U_{oA'} \left(|\Psi''\rangle \otimes |2,0\rangle_{A'} \right) \tag{9.9}$$

is proportional to

$$\alpha \, \langle {}^{j}_{m} {}^{2}_{2} | {}^{j+\sigma}_{m+2} \rangle |j+\sigma, m+2\rangle_{AiB} |1,-1\rangle_{o} |2,-2\rangle_{A'}$$
$$- \beta \, \langle {}^{j}_{m} {}^{2}_{-2} | {}^{j+\sigma}_{m-2} \rangle |j+\sigma, m-2\rangle_{AiB} |1,1\rangle_{o} |2,2\rangle_{A'}. \tag{9.10}$$

Next, Alice tosses her ancilla into the black hole and then measures the black hole's total angular momentum. The $AiBA'$ terms will consist of linear combinations of $|j+\sigma+2, m\rangle, \ \ldots, \ |j+\sigma-2, m\rangle$ weighted by the appropriate Clebsch-Gordan coefficients. If Alice finds $AiBA'$ in a total angular momentum $j+\sigma+\tau$ state, where $\tau \in \{-2, \ldots, 2\}$, it is straightforward to show that the spin that she still holds collapses to the state

$$|\phi''\rangle_{o} \propto \ \alpha \, \langle {}^{j}_{m} {}^{2}_{2} | {}^{j+\sigma}_{m+2} \rangle \langle {}^{j+\sigma}_{m+2} {}^{2}_{-2} | {}^{j+\sigma+\tau}_{m} \rangle |1,-1\rangle_{o}$$
$$- \beta \, \langle {}^{j}_{m} {}^{2}_{-2} | {}^{j+\sigma}_{m-2} \rangle \langle {}^{j+\sigma}_{m-2} {}^{2}_{2} | {}^{j+\sigma+\tau}_{m} \rangle |1,1\rangle_{o}. \tag{9.11}$$

As long as Alice measured the black hole angular momentum at the beginning of the protocol and ensured that $|m| \ll j$, then none of these coefficients vanish. Alice then performs the appropriate unitary transformation on the spin that she holds to restore the state $|\phi\rangle_{o}$.

9.3 Discussion

We now consider several aspects of the proposed algorithm, as well as its consequences for black hole information theory.

9.3.1 State of the Hawking Photons

To see why the Hawking particles must be created in a zero total angular momentum state, note that spacetime is locally flat on the horizon and becomes increasingly flat as the black hole mass M increases. As a result, the only way for a Hawking pair to have non-zero angular momentum is for the pair to pick it up via interactions with the vacuum, i.e., with another Hawking pair. This requires, roughly speaking, that two Hawking pairs be present within one wavelength λ of one another in the time t it takes for a pair to separate. The relevant scaling relations in general are $\lambda \propto T^{-1}$,

$t \propto \lambda$, and $F \propto T^d$, where d is the number of spatial dimensions, T is the Hawking temperature, and F is the particle number flux across the horizon. The fraction f of Hawking pairs which interact with an additional Hawking pair scales at tree order as $f \propto |\mathcal{A}|^2 (F\lambda^{d-1}t)^2 \propto |\mathcal{A}|^2$, where the mass-dependence of the phase-space factors dropped out.[2]

For photons, which are the exponentially dominant form of Hawking radiation at large M, the matrix element $|\mathcal{A}|^2$ must depend on the probability of producing a virtual electron-positron pair to mediate the Hawking pair interaction. This scales as $e^{-m_e/E_\gamma} \sim e^{-m_e GM}$. Thus for large black holes, we expect these interactions to be exceedingly rare, and hence are justified in assuming that the photon pair carries no net angular momentum. We note that the creation of Hawking pairs in the zero angular momentum state relies on the assumption that the local spacetime around the horizon of the black hole is a low-energy, quiescent environment. Were there instead an energetic firewall at the horizon, we could not expect outgoing quanta to come from such a state.

When performing this analysis for other quantum numbers the same arguments apply: for large black holes, the Hawking pair must be created with zero net quantum number. The algorithm we describe will work for any conserved quantum number which photons may carry, so long as the evolution of the relevant sector of the Hilbert space is unitary. Notably, the algorithm does not require the hole's evolution in the total Hilbert space to be unitary over long timescales If the relevant number is not quantized, the information recovered is only up to a precision limit given by the number of bits recovered. For those quantum numbers which photons do not carry, superpositions of states cannot be recovered except by waiting exponentially long in M for the relevant particles to be emitted. If, on the other hand, it is known that a quantum number eigenstate fell in, and hence that only classical information was encoded in this way, then direct measurement of the black hole allows for recovery.

[2] This is not entirely unexpected. Consider, for instance, that the characteristic wavelength of Hawking photons is on the order of the Schwarzschild radius. Roughly speaking, since $t \propto \lambda$, any two photons at the black hole horizon will therefore overlap before they separate.

For example, in order to learn the mass of a particle that fell into the black hole, then one may of course measure the mass of the black hole afterwards, assuming that the initial mass of the black hole was known. Altogether, this allows for unique recovery of classical information about any particle that fell in. This is because each known fundamental particle has a unique set of gauge quantum numbers—mass, spin, charge, and color. This feature is not necessary—it would not hold in a theory with two unbroken $U(1)$ symmetries—but it does hold true in the Standard Model.

9.3.2 Resource Considerations

In its essence, our protocol amounts to a quantum teleportation scheme [240] between a transmitting party—the black hole—and a receiving party—Alice. Its perfect fidelity when $m = 0$ is due to the fact that setting $m = 0$ eliminates any degeneracy in the states that the transmitting party could find after measuring in the total angular momentum basis, as opposed to a (nondegenerate) maximally-entangled basis. Alice would not be able to use an analogous procedure to recover more that a single qubit at a time, since the degeneracy of total angular momentum states rapidly increases as more and more spins are added.

We can also understand the difficulty of the multiple qubit case from the point of view of resources. Suppose that Alice wishes to recover more than a single qubit at a time through a quantum number conservation protocol. As these protocols amount to quantum teleportation schemes, Alice is bound by the resource inequality [241]

$$2[c \to c] + [qq] \geq [q \to q], \tag{9.12}$$

which says that two classical bits, or cbits, of communication and one entangled qubit pair shared between the two parties is necessary to achieve one qubit of communication. If Alice drops N photons into the black hole and collects N Hawking photons, she only obtains $\sim \log_2(N^2) = 2\log_2 N$ cbits since there are $4N + 1$ possible outcomes for the total angular momentum measurement and $\sim 2N$ possible outcomes for the measurement of the projection of the angular momentum along the axis of quantization. As such, she cannot hope to recover some general state of

N qubits, which would require $2N$ cbits. On the other hand, she may be able to recover a state that is encoded in some subspace of $\mathcal{H}$. For instance, Alice could try encoding her data in the total angular momentum of a set of N qubits with total angular momentum s. Thus she is encoding her data in a Hilbert space $\mathcal{H}_s$ with $\dim \mathcal{H}_s = 2s + 1$. Resource considerations do not prohibit the recovery of a state in $\mathcal{H}_s$, which only requires the extraction of $\log_2 \dim \mathcal{H}_s \leq \log_2(N + 1)$ qubits and hence $\sim 2 \log_2 N$ cbits. We suspect that the general method for doing this is similar to the single qubit case.

9.3.3 Timescale Considerations

During the protocol, Alice must wait for the black hole to emit a quantum of Hawking radiation. Hawking emission rates have been calculated by Page [230]; for instance, photons are emitted in their lowest angular momentum mode at a rate given by $t_h^{-1} = 1.463 \times 10^{-4}\ c^3/GM$ for Schwarzschild black holes. Photon emission rates vary as a function of the black hole spin and can be on the order of one hundred times larger in the case of an extremal Kerr black hole [231], so let us express the timescale of Hawking emissions as $t_h = f \cdot GM/c^3$. The factor f contains both geometric and tunnelling factors, and is a function only of the spin of the black hole.

It is interesting to compare the emission time to the scrambling time [53,238,242,243], which may be thought of as the time it takes for Alice's infalling qubit to become incorporated into (the stretched horizon of) the black hole [23]. The scrambling time is

$$t_s = \frac{1}{2\pi T} \ln S, \tag{9.13}$$

where S denotes the entropy of the black hole and where we have used units in which $\hbar$, c, and k_B are 1. This increases faster than t_h as a function of the black hole radius R, since $S \propto R^2$ and $T \propto 1/R$, so there is a critical radius R_{crit} above which the scrambling time is greater than the time required for a Hawking particle to be emitted. In light of our single-qubit protocol, $R > R_{\text{crit}}$ means that the qubit which falls in is essentially bounced off of the black hole, rather than being incorporated

into it. The numerical factors involved, as well as the difference in scaling being in a logarithm, mean that the critical radius for a Schwarzschild black hole is very large ($R_{\text{crit}} \approx e^{853} l_p$, which is considerably larger than the current Hubble radius). However, the dependence of T, S, and the numerical factors on spin means that this radius can be made arbitrarily small by tuning the angular momentum J of the hole, since an extremal Kerr black hole has zero temperature but finite entropy.

9.4 Conclusion

We have described a protocol, based on quantum teleportation, that allows an external observer to recover a single spin qubit that has been dropped into a black hole, if the spin of the hole is measured before and after the qubit is dropped. Our procedure relies on the fact that the angular momentum states of the black hole span the possible states of the qubit; for more than one qubit, this condition would not hold, and an analogous procedure would be unable to recover the information. On the other hand, the fact that an external observer would see apparent information loss due to angular momentum state degeneracy is perhaps interesting in its own right.

This protocol retrieves a very specific kind of information: a single qubit encoded in a conserved quantity such as angular momentum; this is broad enough to include the information contained in any one particle within the Standard Model. Importantly, it is the full quantum state of the qubit, not merely the classical angular momentum. While our protocol does not extend to information encoded in the entanglement between multiple particles, the general idea of using quantum teleportation to recover information deserves further study.

Acknowledgements

We would like to thank Gil Refael, who was instrumental in the conception of this protocol. We also thank Ning Bao, Achim Kempf, Stefan Leichenauer, Don Marolf, Don Page, John Preskill, Grant Remmen, and Guillaume Verdon-Akzam for helpful

discussions. This research was funded in part by the Walter Burke Institute for Theoretical Physics at Caltech, by DOE grant DE-SC0011632, and by the Gordon and Betty Moore Foundation through Grant 776 to the Caltech Moore Center for Theoretical Cosmology and Physics. ACD was supported by the NSERC Postgraduate Scholarship program. ASJ was supported by a Barry M. Goldwater Scholarship.

Chapter 10

RESCUING COMPLEMENTARITY WITH LITTLE DRAMA

The AMPS paradox challenges black hole complementarity by apparently constructing a way for an observer to bring information from the outside of the black hole into its interior if there is no drama at its horizon, making manifest a violation of monogamy of entanglement. We propose a new resolution to the paradox: this violation cannot be explicitly checked by an infalling observer in the finite proper time they have to live after crossing the horizon. Our resolution depends on a weak relaxation of the no-drama condition (we call it "little-drama") which is the "complementarity dual" of scrambling of information on the stretched horizon. When translated to the description of the black hole interior, this implies that the fine-grained quantum information of infalling matter is rapidly diffused across the entire interior while classical observables and coarse-grained geometry remain unaffected. Under the assumption that information has diffused throughout the interior, we consider the difficulty of the information-theoretic task that an observer must perform after crossing the event horizon of a Schwarzschild black hole in order to verify a violation of monogamy of entanglement. We find that the time required to complete a necessary subroutine of this task, namely the decoding of Bell pairs from the interior and the late radiation, takes longer than the maximum amount of time that an observer can spend inside the black hole before hitting the singularity. Therefore, an infalling observer cannot observe monogamy violation before encountering the singularity.

This chapter was published as Ref. [8], N. Bao, A. Bouland, A. Chatwin-Davies, J. Pollack, and H. Yuen, "Rescuing complementarity with little drama," JHEP ***12*** *(2016) 026,* `arXiv:1607.05141`.

10.1 Introduction

The information paradox [233] and its more modern AMPS incarnation [18, 19] are deeply puzzling issues lying at the center of any attempts at reconciling quantum mechanics with gravity. Black hole complementarity, as proposed by [23], attempted to resolve the information paradox by asserting that information that falls into the black hole interior is also retained at the stretched horizon. Observers are only able to access this information in one of two "complementary" descriptions, either in the interior or at the horizon, so that the apparent violation of the no-cloning theorem visible in a global description could never be verified. AMPS, however, considered a scenario in which an observer first collects information on the outside by gathering Hawking radiation, then jumps through the horizon and into the black hole interior. Assuming standard postulates of black hole complementarity, namely

1. unitarity,
2. the validity of low-energy effective field theory outside the stretched horizon,
3. that the black hole is a quantum mechanical system with dimension given by $e^{A/4}$,

and further

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

AMPS pointed out an apparent violation of monogamy of entanglement[1] among three systems: the black hole interior, the recently emitted Hawking radiation (late radiation), and the previously emitted Hawking radiation (early radiation). To avoid this violation, it therefore seemed necessary to give up one of the assumptions mentioned above, all of which are cherished pillars of modern physics. Giving up the

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

final assumption would mean that observers who attempt to enter the black hole would be violently destroyed by high-energy excitations, hence the name "firewall paradox."

This led to a flurry of attempts to resolve the paradox by weakening one or more of the core axioms, or by changing the paradigm completely [174, 244–252]. Reaching consensus as to which resolution is the correct one has proven challenging.

An interesting proposed resolution to the information paradox, based on arguments from computational complexity, was given by Harlow and Hayden [253]. They argued that the part of the AMPS experiment where the experimenter has to decode[2] entanglement between the old radiation and the late radiation of the black hole involves an extremely difficult computational task. Under very plausible conjectures in computational complexity[3], the time required to perform this quantum computation in general would be exponentially longer than the evaporation time of the black hole. Thus, by the time that the entanglement is decoded, there will remain no black hole within which to check for the violation of monogamy of entanglement. While the two quantum mechanical descriptions of the black hole appear to imply a violation of monogamy, this apparent violation cannot be "revealed" by the AMPS experiment, and thus the experimenter does not see any contradiction with quantum mechanics. Just like the original violation of no-cloning in black hole complementarity itself, this would signal that only the various partial descriptions accessible by a single observer should be considered.

The main appeal of this argument is that it does not require a weakening of any of the core assumptions mentioned previously. However, it is not without its vulnerabilities. For example, Oppenheim and Unruh [255] gave an argument showing that a very motivated experimenter could evade the Harlow-Hayden complexity barrier by offloading the hard computation into a "precomputation" phase before the black

[2]To "decode the entanglement" of a state $|\psi\rangle_{AB}$ is to act with local unitaries on A and B to create a Bell pair across A and B. This is similar to the notion of entanglement distillation [254], except here we have only one copy of the state $|\psi\rangle_{AB}$, whereas in distillation one has multiple identical copies of the state.

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

hole had even formed, and then perform the AMPS experiment efficiently using the "cached computation." Another vulnerability is that the computational hardness of the Harlow-Hayden argument assumes that the black hole in question somehow encodes a cryptographically difficult one-way function; however, one may be able to set up a black hole so that the entanglement decoding task is particularly easy [256].

Nevertheless, the Harlow-Hayden proposal remains a compelling one, and it sets the context for the argument that we present in this paper. Here, we also study whether ideas from information theory and computer science can help resolve the information paradox, but in another setting: whereas Harlow and Hayden focus on the computational complexity of the AMPS experiment *outside* the black hole, we examine the information processing that must be performed *inside* the black hole in order to check for violations of monogamy of entanglement. This is a potentially different line of argument, because while it might be possible to evade computational limits outside of the horizon [19, 255], one certainly cannot extend one's time inside the horizon, as an infalling observer invariably hits the singularity in a bounded amount of time.

In this paper we study an observer who begins outside of an evaporating Schwarzschild black hole well after the Page time and who has learned that a subset of late Hawking radiation that she holds is maximally entangled with the early Hawking radiation[4]. We suppose that the observer then enters the black hole, sees no firewall, and then attempts to decode maximal entanglement between the late radiation that she holds and the black hole interior. If she succeeds in completing this task, she can then perform measurements on an ensemble of her decoded Bell pairs in order to probabilistically detect a violation of monogamy of entanglement. We compare the proper time it takes for the observer to perform this procedure with the infall time before the observer hits the singularity. We find that, under the assumption that the subsystem of the black hole interior with which the observer's late radiation is entangled has diffused throughout the whole interior at the time she crosses the horizon, the

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

observer will not have enough time to complete even the first step of the procedure, i.e., entanglement decoding, before encountering the singularity. As such, while a global description, if it existed, would contain an implicit violation of monogamy of entanglement, an observer who entered the black hole would unable to directly verify any such violation. Therefore, our resolution of the firewalls paradox is similar in spirit to complementarity [23] in the sense that apparent global violations of quantum mechanics are not verifiable by local observers.

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

The organization of this paper is as follows. In Sec. 10.2, we review facts about black holes and their scrambling from the perspective of different observers in spacetime. In Sec. 10.3, we focus on the specific task of collecting a late-time Hawking radiation particle, assess the degree of scrambling that has occurred prior to the observer crossing the stretched horizon of the black hole, and give a discussion of the little-drama condition. In Sec. 10.4, we combine all the ingredients from the previous sections and analyze the time needed to perform the task of checking for violations of monogamy. Finally, we discuss and conclude in Sec. 10.5 and Sec. 10.6.

10.2 Background: Black Holes and Scrambling

In the thought experiments to follow, we will consider black holes that formed from the gravitational collapse of matter and that eventually evaporate into a gas of Hawking radiation. We will assume that the initial mass of any black hole that we consider is large enough that physics outside the black hole is well-described

by effective field theory on a black hole background in regions of spacetime that are sufficiently distant from the end of evaporation. We will also suppose that the process of black hole formation and evaporation is a fundamentally unitary process. As such, if the matter that collapsed to form a black hole was initially in a pure quantum state, then the state of the Hawking radiation after evaporation—as well as any combined intermediate state of the black hole and hitherto emitted Hawking radiation—is also a pure state.

Consider now some observer who resides outside the black hole. We will adopt the viewpoint that such an observer's observations are described according to complementarity [23] and the membrane paradigm [133]. Explicitly, suppose that the black hole spacetime is foliated by some set of achronal (spacelike or null) surfaces with respect to which the observer performs field-theoretic calculations. In accordance with complementarity, an observer outside the black hole should not associate a Hilbert space to an entire surface Σ if it intersects the event horizon. In such a case, she instead organizes the physical Hilbert space associated to Σ into a tensor product $\mathcal{H} = O \otimes D$. The space O describes the degrees of freedom on the portion of Σ that lies outside of the black hole, and D is a Hilbert space that describes the black hole's degrees of freedom and that is localized about the event horizon (Fig. 10.1). From the outside observer's point of view, all of physics is described by, and all processes play out in, these two Hilbert spaces; she never has to (and in fact *may not*) make reference to the the black hole interior.[5]

We will suppose that D is localized to the stretched horizon of the black hole [23]. We take the outer boundary of the stretched horizon to be at a proper distance on the order of a Planck length above the event horizon. As such, the outer boundary of the stretched horizon is a timelike surface with which an outside observer can interact.

Despite the fact that a complete theory of quantum gravity is not known and that

[5]See also [113] (in particular Sec. 4) as well as Sec. 10.5.4 for further discussion of the way in which $\mathcal{H}$ factorizes and the ways in which different factorizations are related as a consequence of assuming complementarity.

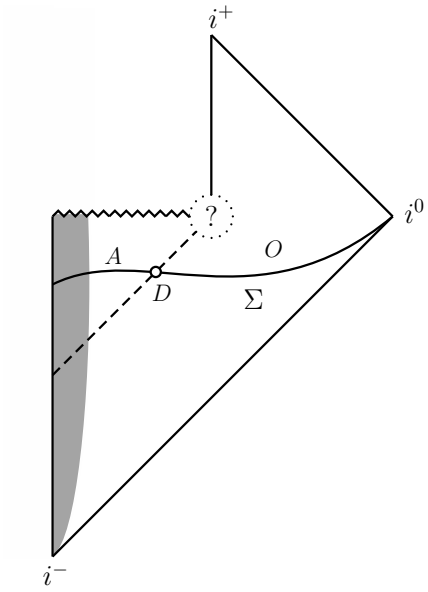

Figure 10.1: Penrose diagram of a black hole that forms from the gravitational collapse of matter and that ultimately evaporates.

the full dynamics of black holes are not understood, it is widely expected that the quantum state of matter gets scrambled when it enters the stretched horizon [42,257, 258]. There are many possible ways to define scrambling, but informally speaking, a system scrambles if it diffuses quantum information over all its degrees of freedom. In particular, a black hole has scrambled the information in a small subset $D' \subset D$ when any initial entanglement between D' and the outside O gets distributed evenly throughout D, i.e., when almost all small subsets of D have nearly the same amount of entanglement with O. After scrambling, an observer cannot recover this entanglement unless she examines a sizable fraction of the entire horizon D.

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

$$t_s = \frac{1}{2\pi T} \ln S \,, \tag{10.1}$$

where T and S are the temperature and entropy of the black hole respectively [42, 53, 242, 243, 257]. (Both in this expression and throughout the paper we have set $c = k_{\rm B} = \hbar = 1$.) This time is measured with respect to the clock of an asymptotic observer who is far away from the black hole. For example, for a Schwarzschild black

hole in $3+1$ dimensions, the metric is given by

$$ds^2 = -\left(1 - \frac{r_s}{r}\right) dt^2 + \left(1 - \frac{r_s}{r}\right)^{-1} dr^2 + r^2\, d\Omega_2^2, \tag{10.2}$$

the temperature is

$$T = \frac{1}{8\pi GM} = \frac{1}{4\pi r_s}, \tag{10.3}$$

and the entropy is

$$S = \frac{A}{4G} = \frac{4\pi {r_s}^2}{4l_P^2} = \frac{\pi {r_s}^2}{l_P^2}. \tag{10.4}$$

As such, the scrambling time is given by

$$t_s = r_s \ln \frac{\sqrt{\pi} r_s}{l_P}\,. \tag{10.5}$$

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

$$\tau_s(r_0) = \sqrt{1 - \frac{r_s}{r_0}}\, t_s\,. \tag{10.6}$$

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

$$\delta r = \frac{l_P^2}{4r_s} + O\left(\frac{l_P^3}{{r_s}^2}\right)\,. \tag{10.7}$$

It then follows that

$$\begin{aligned} \tau_s(r_s + \delta r) &= \sqrt{\frac{l_P^2}{l_P^2 + 4{r_s}^2}}\, r_s \ln\left[\frac{\sqrt{\pi} r_s}{l_P}\right] \\ &\approx \frac{l_P}{2} \ln\left[\frac{\sqrt{\pi} r_s}{l_P}\right]\,, \end{aligned} \tag{10.8}$$

which is consistent with other calculations of the scrambling time at the stretched horizon [42, 257].

10.3 Hawking radiation and scrambling: what Alice sees

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

10.3.1 Scrambling, inside and out

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

Recall that Hawking radiation consists of paired entangled excitations of field modes. The outgoing modes constitute the radiation that is visible to stationary observers,

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

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

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

To be more precise, suppose that Alice follows a timelike trajectory $\mathcal{A}$ that crosses the event horizon and ultimately hits the singularity, as shown in Fig. 10.2. (Partially) foliate the spacetime with her past lightcones. When she is inside the black hole, we associate A to the portion of her lightcone that lies inside the black hole. For all of her lightcones, we associate O to the part of the lightcone that lies outside the

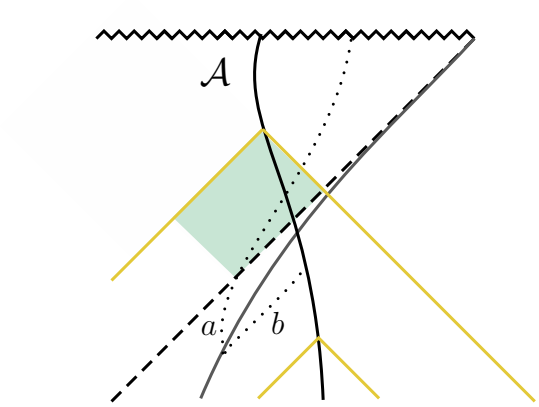

Figure 10.2: Alice's trajectory $\mathcal{A}$ and past lightcones (shown in yellow) as she falls toward the singularity. The stretched horizon is shown in grey, and the trajectories of the outgoing and ingoing Hawking particles are shown as dotted lines. We suggest that scrambling causes information about the ingoing excitation to spread out behind the event horizon so that it is delocalized on the intersection of Alice's past lightcones with the causal future of the excitation's horizon crossing point (shaded region).

black hole and D to the surface where her lightcone intersects the stretched horizon. According to complementarity, we postulate that for each lightcone whose tip lies inside the black hole, there exists a unitary map

$$\mathcal{U}_{\text{comp}} \,:\, D \otimes O \;\longrightarrow\; A \otimes O \tag{10.9}$$

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

$$\tilde{U}_{\text{scr}} \;\equiv\; \mathcal{U}_{\text{comp}}\,(U_{\text{scr}} \otimes I_{\text{out}})\,\mathcal{U}_{\text{comp}}^{\dagger}\,. \tag{10.10}$$

Intuitively, one would expect that scrambling should persist behind the event horizon. For instance, if one were to drop a qubit into the stretched horizon and wait for it to be well-scrambled, it would be surprising to find it more or less intact and localized after jumping into the black hole. Moreover, such a discovery would be troubling in light of Hayden and Preskill's finding that the information contained in that qubit is very rapidly returned to the exterior of the black hole [42]. Mathematically, this expectation is equivalent to the statement that we do not expect

the unitary operator (10.10) to act trivially on the physically relevant states in A. We note, however, that it is not logically impossible that $\mathcal{U}_{\text{comp}}$ exactly undoes the action of U_{scr}.

On the other hand, it would also be desirable to reconcile the unitary (10.10) with the semiclassical expectation that spacetime and macroscopic gravitating objects near the event horizon are well-described by general relativity. Put another way, the field equations of general relativity should be sufficient, at least to a first approximation, to track classical matter thrown into the black hole on timescales where Hawking evaporation is unimportant. For example, from a semiclassical point of view, if you were to drop a rock into a black hole, you would still expect to find the rock on its freefall trajectory if you accelerated to catch up with it behind the event horizon.

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

We emphasize that the dynamics that we have proposed constitute a violation of

[6]See Sec. 10.5 for a discussion on this simplifying assumption.

the no-drama condition, albeit a far milder one than firewalls. In classical general relativity, the equivalence principle remains intact: the black hole geometry is still described by the Schwarzschild metric, and nothing special happens at the horizon. Even semiclassically, expectation values of operators should remain unchanged: we are not changing the emission rate of Hawking quanta or the effective temperature of the black hole. However, working with Hawking emission on a particle-by-particle basis requires a more detailed description. We can write the quantum state describing the evaporating black hole in a basis of states which each contain Hawking particles. In each basis state, individual Hawking quanta are pair-produced as genuine particles (i.e., wavepackets) at a specific spot on the horizon of the black hole, with one wavepacket excitation describing a particle produced in A and a corresponding particle in O. In each basis state, $\tilde{U}_{\text{scr}}$ acts to rapidly spread the excitation in A into many other modes, so that after a scrambling time it can no longer be described as a wavepacket or particle. It is this evolution, which differs dramatically from the propagation of a particle on an empty background metric, that can be seen as violating no drama.

10.3.2 Scrambling and kinematics

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

In principle, Alice could wait arbitrarily closely to the stretched horizon so that the ingoing excitation has little time to scramble. We note, however, that the scrambling time at the stretched horizon is a fantastically small amount of time. For example, for a supermassive black hole like Sagittarius A* with a mass of about four million solar masses, Eq. (10.8) predicts that the scrambling time at the stretched horizon

should be $\tau_s \approx 3 \times 10^{-42}$ s, or about 50 Planck times. As such, Alice does not have much time at all outside of the black hole before scrambling happens, and in practice she will have some amount of computational overhead if she verifies the entanglement between late radiation and early radiation before entering the black hole. Furthermore, if Alice collects $k > 1$ Hawking particles, then scrambling of the first $k-1$ ingoing excitations is virtually guaranteed to have happened before Alice can cross the horizon. This is because the average rate of Hawking emissions is (much) slower than the rate of scrambling [230, 231]. Consequently, instances where Alice can cross the horizon before ingoing modes have scrambled are $(k-1)$-fold exponentially suppressed.[7] As we will discuss in the next section, Alice will need to collect $k > 1$ Hawking particles in order to be statistically confident in her measurements inside the horizon.

Separately from the considerations above, it is also interesting to ask what the theoretical minimum height at which Alice can wait above the black hole is above which scrambling is guaranteed to have happened when Alice enters the black hole. This is the height for which exactly one scrambling time elapses at the stretched horizon in the time it takes a light ray to make a round trip between the stretched horizon and a mirror at the height in question. This situation is depicted in Schwarzschild coordinates in Fig. 10.3.

The radial lightlike geodesics are given by

$$r - (r_s + \delta r) + r_s \ln\left[\frac{r_s - r}{r_s - (r_s + \delta r)}\right] = \pm\left(t + \frac{t_s}{2}\right) , \tag{10.11}$$

with t_s and δr as given in Eqs. (10.5) and (10.7) respectively. The minimum coordinate height is obtained by setting $t = 0$ in Eq. (10.11) and solving for r:

$$r_{\min} = r_s \cdot W\left(\frac{\delta r}{r_s} \exp\left[\frac{2\delta r + t_s}{2r_s}\right]\right) . \tag{10.12}$$

[7] From [230], the cumulative Hawking emission rate for a Schwarzschild black hole is about $10^{-4}\ c^3/GM$, so take the characteristic timescale of Hawking emissions to be $t_H \sim 10^4\ GM/c^3$. Note that this is measured in Schwarzschild time, so with the relevant boost factor of $l_P/2r_s$ and for the supermassive black hole discussed above, the characteristic (proper) timescale of Hawking emissions at the stretched horizon is about $(10^3 - 10^4)\ l_P/c$, which is much larger than the scrambling time. Also c.f. footnote 9 below.

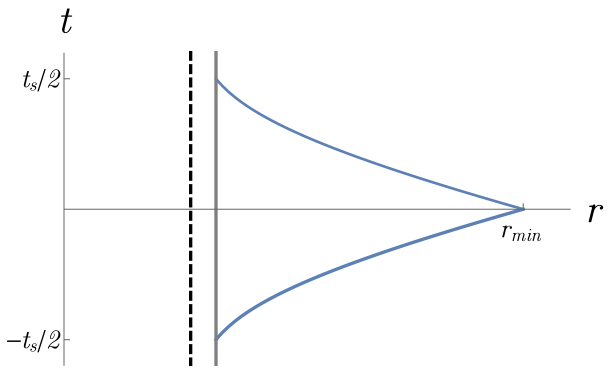

Figure 10.3: Minimum height above which scrambling is guaranteed to occur.

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

$$\begin{aligned}\tilde{r}_{\min} &= \int_{r_s}^{r_{\min}} \left(1 - \frac{r_s}{r}\right)^{-1} dr \\ &= 2\sqrt{r_s e^{t_s/2R}}\sqrt{\delta r} + O\left((\delta r)^{3/2}\right) \\ &\approx \sqrt{\pi} r_s \,. \end{aligned} \tag{10.13}$$

This result is interesting in light of proposals by Nomura, Sanches, and Weinberg [251] and by Giddings [259] which both suggest that Hawking radiation is largely invisible to observers unless they are at least on the order of a few Schwarzschild radii away from the horizon of a black hole, which further limits Alice's ability to evade scrambling.

10.4 Computation behind the horizon

To summarize the previous section, if excitations at the stretched horizon are scrambled when Alice reaches the stretched horizon, then we are proposing that the state of the ingoing Hawking modes is thoroughly mixed with other degrees of freedom in the black hole's interior. In this section we assume that this scrambling has had time to occur; as we explain in Sec. 10.3.2, such a situation should be generic. As such, Alice is forced to access and process a large number of degrees of freedom that are distributed throughout the interior of the black hole if she wants to verify monogamy of entanglement. In this section, we discuss how to model the task of

verifying entanglement and we investigate its complexity. In the rest of this paper we will set $l_P = 1$ for brevity.

10.4.1 Model for verifying entanglement

Following the convention of [18], we continue to denote the Hilbert space of the interior of the black hole by A, and we label the Hilbert spaces of the early radiation and late radiation by R and B respectively (so that R and B are subsets of the space O that we defined in Sec. 10.2). Let $b^{(k)} \subset B$ denote the Hilbert space of the k outgoing Hawking modes that Alice collected and $a^{(k)} \subset A$ the Hilbert space of the corresponding k ingoing modes. We model $b^{(k)}$ and $a^{(k)}$ each as a collection of k qubits. Referring to Eq. (10.9), since the Hilbert space O is the same in both complementary descriptions of physics[8], it follows that $|A| = |D| = e^{S_{BH}}$, where S_{BH} is the Bekenstein-Hawking entropy of the black hole and where $|\cdot|$ denotes the dimension of a Hilbert space. As such, we model A as a collection of $n \sim S_{BH}$ qubits that are distributed throughout the interior of the black hole and that are visible to Alice on her past lightcones.

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

What do we mean by verifying entanglement? In quantum theory, there is no

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

measurement that reliably distinguishes between entangled states and unentangled states—this is because the set of unentangled pure states is non-convex. However, it is possible to *statistically* test if an unknown state is in a *particular* entangled state. For example, if we let $|\Phi\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ denote an EPR pair, then the two-outcome measurement $M = \{|\Phi\rangle\langle\Phi|, I - |\Phi\rangle\langle\Phi|\}$ will probabilistically indicate whether a given pair of particles $|\psi\rangle$ is an EPR pair or not. If $|\psi\rangle$ is indeed an EPR pair, then this measurement will always return outcome $|\Phi\rangle\langle\Phi|$ with certainty. On the other hand, if $|\psi\rangle$ is an unentangled state $|\phi\rangle \otimes |\theta\rangle$, then it will return outcome $I - |\Phi\rangle\langle\Phi|$ with probability at least 1/2. While the error of this statistical test is rather large, it can be reduced exponentially by repeating it many times. Let V and W denote two disjointed quantum systems. When we say that Alice has "verified maximal entanglement between V and W," we mean that Alice has decoded k pairs of particles from V and W, measured each pair using the two outcome measurement M, and verified that all k pairs projected to an EPR pair. This occurs with probability 1 if Alice did indeed decode k EPR pairs; if V and W were unentangled, then this occurs with probability at most 2^{-k}. Therefore as k grows, the probability that Alice thinks that V and W are entangled (when they are actually unentangled) becomes exponentially small. For example, if Alice wants to obtain 5 sigma certainty (error probability 1 in 3.5 million) that V and W share maximally entangled particles, she only needs to decode $k = 22$ EPR pairs from V and W.

10.4.2 Alice's computational task

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

$b^{(k)}a^{(k)}$ were produced[9]. The state of the Hawking pairs and the black hole interior can be described by the density matrix

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

where $|\Phi\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ is a maximally entangled Hawking pair, and ρ^A is the density matrix of the black hole interior right before the pair production event. By Page's theorem [40, 41], after the Page time ρ^A is close to being maximally mixed; for the remainder of this argument, we will assume that ρ^A is exactly the maximally mixed state on n qubits.[10]

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

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

Because our understanding of the quantum mechanical evolution of black holes is rather limited, we will model the unitary U as being Haar-random. (In fact our arguments will carry through in the case that U is chosen from an ensemble of efficiently constructible unitaries that is sufficiently randomizing; we will discuss this in more detail in Sec. 10.5.)

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

[9] For simplicity here we assume that they are produced simultaneously rather than one-by-one, but this does not hinder the argument. Indeed, if they are produced sequentially, then due to arguments by Page [230, 231], the average rate of Hawking pair production is less than one pair per scrambling time. Therefore, in a sequential production picture, all but the last Hawking pair will have been scrambled by the time that Alice can enter the black hole. If Hawking radiation can be modeled thermally, sequential emission is exponentially preferred over simultaneous emission.

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

Then, by [42], we have that

$$\int dU \left\| \tau^{b^{(k)}S} - \tau^{b^{(k)}} \otimes \tau^{S} \right\|_1^2 \leq |b^{(k)}S| \cdot \mathrm{Tr}\left[(\sigma^{b^{(k)}A'})^2 \right]. \tag{10.14}$$

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

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

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

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

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

10.5 Discussion

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

10.5.1 Modeling scrambling dynamics

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

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

As noted by Hayden and Preskill [42], one can model the dynamics of a black hole using *random unitary designs*. Informally speaking, unitary designs are ensembles of unitaries with polynomial circuit complexity that in many respects behave like Haar-random unitaries. In our argument, the Haar unitary ensemble can be replaced by an (approximate) unitary design and our conclusion remains essentially unchanged: unitary designs, though possessing small circuit complexity, still "smear" quantum information across all degrees of freedom. Unitary designs have been extensively

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

studied in the quantum information community. By now, we know several examples of (approximate) unitary designs [260, 261].

Still, what do we mean when we say that a *particular* black hole behaves like a unitary randomly chosen from an ensemble? After all, a black hole behaves according to none other but *the* unitary given by the theory of quantum gravity. Unfortunately, since this theory is still unavailable to us, in our calculations we must make a "best guess" at what a black hole unitary must look like. Without presupposing unjustified constraints on the theory of quantum gravity, our best guess for black hole dynamics is that the Hamiltonian governing the interior should be local and strongly mixing, and that the black hole evolves in polynomial time. The Maximum Entropy Principle from statistics and learning theory tells us that our best guess for the black hole unitary is a randomly chosen one from the uniform distribution over unitaries with polynomial circuit complexity[13]. We note that this ensemble of unitaries is known to form an approximate unitary design [260], and thus has the scrambling properties required by our argument.

10.5.2 Black holes in other dimensions.

One may also object that this argument is specific to spacetimes of dimension 3+1. In higher dimensions this argument only becomes stronger, since in spacetimes with spatial dimension d, the number of qubits that make up the interior Hilbert space, $|A|$, scales like $O(r_s^{d-2})$, while the infall time scales like $O(r_s)$. As such, the infall time is increasingly smaller with respect to $|A|$ for $d > 3$. But, this is not necessarily true for lower spatial dimensions. For example, in AdS_3, the number of qubits and the infall time both scale linearly with r_s. Consequently, our previous trivial bound on the number of accessible qubits does not suffice here. In this case one can appeal to the fast scrambling conjecture to render the computation impossible. The fast scrambling conjecture of Sekino and Susskind [257] states (among other things) that

[13]The Maximum Entropy Principle is a formalization of Occam's Razor in machine learning and statistical learning theory [262]. It says that, given a set of hypotheses consistent with one's observations, one's best hypothesis is the maximum entropy one: a randomly chosen one from that set.

black holes are the fastest scramblers in nature[14]. Lashkari *et al.* [258] formalized this notion in terms of quantum information by stating that black holes saturate the $r_s \log r_s$ lower bound for scrambling time. In this work, we consider a quantum complexity formulation:

Conjecture 10.5.1. *Let $k \ll n/2$, i.e., let k be much smaller than the number of qubits in the black hole. Let U be the unitary corresponding to running black hole dynamics for time t on $A' = a^{(k)}A$, as measured by an asymptotic observer. Then recovering the entanglement between $a^{(k)}$ and $b^{(k)}$ from A' and $b^{(k)}$ requires time at least t. More formally, for any unitary V acting on system A', if $\nu^{b^{(k)}A'} = (I^{b^{(k)}} \otimes V^{A'}U^{A'})\sigma^{b^{(k)}A'}(I^{b^{(k)}} \otimes V^{A'}U^{A'})^\dagger$ is the state of the system after applying VU to A' , and if*

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

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

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

10.5.3 Localization of the experimenter.

In our argument, we assume that Alice is localized throughout our experiment, and therefore can access only $O(r_s)$ qubits after crossing the horizon. One might object that if one knew the exact dynamics of $\tilde{U}_{\text{scr}}$, one could set up the infalling matter such that a nonlocal experiment is performed on the interior modes and the result is then sent to Alice. However, this is impossible because Alice is out of causal contact

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

with most of the black hole interior [263] from which the results of the nonlocal experiment would have to be sent. Therefore, even this non-local experiment cannot reveal entanglement between the interior and exterior Hawking modes before Alice hits the singularity.

10.5.4 Relation to prior works

We first note that in [263] arguments have already been made about the inability of the infalling observer to access the entirety of the interior of the black hole except at the singularity. These arguments are quite different in nature from the information-theoretic ones of this paper. In particular, there appears to be the possibility to work around the arguments in [263] by using multiple observers [264], something which does not seem to be an issue in the more information-theoretic arguments of this note.

Readers may notice that our argument significantly resembles that given by Hayden and Preskill [42]. While the techniques are similar, our conclusions and assumptions differ in several ways. First, [42] concludes that black holes, rather than being information sinks, are plausibly more like information "mirrors;" information deposited into the black hole gets released (in scrambled form) as quickly as possible. On the other hand, our goal is to demonstrate a *lower bound* on Alice's ability to recover a single qubit of information within the black hole after it has been scrambled. Second, Hayden and Preskill explicitly model the joint state of the black hole, its radiation, as well as some reference system as a pure state. However, in the context of the firewalls paradox, we cannot write down such a description to begin with! In our setting, we focus solely on the part of the black hole that Alice sees after she has collected her Hawking mode and has crossed the event horizon. This is consistent with complementarity; we only need to provide a valid description of physics inside the horizon, which need not be in a tensor product with the description of physics outside the horizon.

Our proposal also shares some spiritual similarities with fuzzball complementarity

[265], in which undisturbed freefall through the horizon is recovered in the limit where the incident energy of the observer is much larger than the temperature of the black hole, in the sense that local properties of the infalling observer are important to consider in both cases. We note that in the context of the fuzzball program, the definition of complementarity invoked by AMPS—which we follow in Sec. 10.2 when we define the Hilbert space relevant to the problem—is replaced by a different and perhaps more correct definition involving the definition of the state along the complete slice, both inside and outside of the horizon. While it would be interesting to reformulate our results in that lens, it is perhaps unnecessary: in that limit the fuzzballs program already precludes the need for a different resolution to the information paradox! Instead, we emphasize that, even when cleaving as close to AMPS-style complementarity definitions as possible, information- and complexity-theoretic arguments by themselves strongly constrain the ability for any observer to actually observe violation of monogamy of entanglement.

We also differ from the fuzzballs approach in analyzing operationally what is possible for the observer to compute after crossing the stretched horizon of the black hole on the way to an existent singularity. In this work, the singularity plays a vital role in determining the longest possible time available to perform the computation. But, in fuzzball complementarity, the singularity is fuzzed out and resolved at some characteristic fuzzball radius, behind which space stops existing. It may be interesting to see by what degree our bounds would tighten in the specific case of fuzzballs; we reiterate, though, that we are already able to demonstrate that we cannot operationally detect monogamy of entanglement even without the shorter longest possible time for the computation given by the fuzzball program.

Finally, we also note the recent paper [266], which provides a concrete toy model for fuzzball complementarity. It would be interesting to examine our proposals in the context of this work, since the dynamics of infalling excitations discussed in [266] may be able to inspire and inform a similarly concrete realization of the scrambling dynamics that we discussed in Sec. 10.3.1.

10.5.5 Other black hole geometries

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

We note, however, that the inner horizon is not entirely understood, both from the perspective of general relativity and quantum theory [267, 268]. (For example, the inner horizon is strongly believed to be unstable.) As such, it is likely that our assumptions about quantum mechanics and general relativity would need to be modified (at least in the vicinity of the inner horizon) in order to discuss charged spinning black holes, and it is another question entirely what form the AMPS paradox would take if it persists.

10.6 Conclusion

We have described a resolution of the information paradox that amounts to a weakening of the no-drama condition — a new condition that we call little-drama. We suppose that quantum systems that cross the event horizon of a black hole experience nontrivial evolution which entangles them with other degrees of freedom in the black hole interior. Such evolution inside the horizon is the complementary description of scrambling on the stretched horizon and constitutes a mild departure from the predictions of a non-gravitating field theory.

The little-drama condition allows for an apparent violation of monogamy of entanglement that is similar in spirit to the Harlow-Hayden proposal. Past the Page

time, an observer can verify that early and late Hawking radiation have the right entanglement structure outside of a black hole and then smoothly pass through the event horizon. While the smooth crossing implies a violation of monogamy of entanglement—it would seem that the late radiation is maximally entangled with both the early radiation and the black hole interior—we found that the observer could not verify this violation before encountering the singularity.

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

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

Acknowledgements

We thank Wilson Brenna, Charles Cao, Sean Carroll, Daniel Harlow, Jonathan David Maltz, Grant Remmen, and Douglas Stanford for helpful discussions. This is material is based upon work supported in part by the following funding sources: N.B. was supported in part by the DuBridge Postdoctoral Fellowship, by the Institute for Quantum Information and Matter, an NSF Physics Frontiers Center (NFS Grant PHY-1125565) with support of the Gordon and Betty Moore Foundation (GBMF-12500028), and by the U.S. Department of Energy, Office of Science, Office of High Energy Physics, under Award Number DE-SC0011632. A.B. was supported in part by the NSF Graduate Research Fellowship under grant no. 1122374 and by the NSF Alan T. Waterman award under grant no. 1249349. A.C.-D. was supported by the NSERC Postgraduate Scholarship program. J.P was supported in part by DOE grant DE-SC0011632 and by the Gordon and Betty Moore Foundation through Grant 776 to the Caltech Moore Center for Theoretical Cosmology and Physics. H.Y. was supported by Simons Foundation grant 360893, and National Science Foundation grant 1218547.

Chapter 11

BRANCHES OF THE BLACK HOLE WAVE FUNCTION NEED NOT CONTAIN FIREWALLS

We discuss the branching structure of the quantum-gravitational wave function that describes the evaporation of a black hole. A global wave function which initially describes a classical Schwarzschild geometry is continually decohered into distinct semiclassical branches by the emission of Hawking radiation. The laws of quantum mechanics dictate that the wave function evolves unitarily, but this unitary evolution is only manifest when considering the global description of the wave function: it is not implemented by time evolution on a single semiclassical branch. Conversely, geometric notions like the position or smoothness of a horizon only make sense on the level of individual branches. We consider the implications of this picture for probes of black holes by classical observers in definite geometries, like those involved in the AMPS construction. We argue that individual branches can describe semiclassical geometries free of firewalls, even as the global wave function evolves unitarily. We show that the pointer states of infalling detectors that are robust under Hamiltonian evolution are distinct from, and incompatible with, those of exterior detectors stationary with respect to the black hole horizon, in the sense that the pointer bases are related to each other via nontrivial transformations that mix system, apparatus, and environment. This result describes a Hilbert-space version of black hole complementarity.

This chapter is available in preprint form as Ref. [9], N. Bao, S. M. Carroll, A. Chatwin-Davies, J. Pollack, and G. N. Remmen, "Branches of the black hole wave function need not contain firewalls," `arXiv:1712.04955`.

11.1 The black hole information puzzle

In 1975, Stephen Hawking showed that, in coordinates stationary with respect to a static black hole, quantum fields outside the black hole horizon are in a state of outgoing radiation that is very nearly thermal [43]. The backreaction of this thermal emission should lead even astrophysical black holes to evaporate over time, gradually transferring their mass into an ensemble of dilute radiation. However, upon extrapolating Hawking's result to the case of a completely evaporating black hole, one is confronted with an apparent departure from quantum mechanics: it appears that when a pure state of matter—and the quantum information that it encodes—collapses into a black hole that then evaporates, it has evolved into a thermal mixed state and lost its coherent information. Whether and how the quantum state can evolve unitarily from before a black hole is formed to after it evaporates is known as the black hole information puzzle [23, 43, 44, 269].

Several renditions of the black hole information puzzle have emerged over the last few decades. In its modern form, the information puzzle is neatly summarized as a conflict between the following four postulates, articulated by Almheiri et al. (AMPS) [18]:

1. *Unitarity* — As viewed by an observer who remains far away from the black hole, the formation and evaporation of the hole is a unitary quantum-mechanical process.

2. *Local Effective Field Theory* — To the exterior of the black hole's stretched horizon [23, 133], the physics of matter is well described by a local effective field theory on a black hole spacetime background.

3. $S_{\rm bh} = S_{\rm BH}$ — As viewed by an observer who remains far away from the horizon, the black hole is a quantum-mechanical system that is represented by a finite dimensional Hilbert space. Moreover, the von Neumann entropy of an old black hole, $S_{\rm bh}$, is (if not exactly, approximately) equal to its Bekenstein-Hawking entropy, $S_{\rm BH}$.

4. *No Drama* — An observer who crosses the apparent horizon of the black hole (but remains far from its central singularity) encounters nothing that runs contrary to the predictions of semiclassical general relativity and effective field theory.

Taken together, these postulates seemingly imply a violation of monogamy of entanglement [270]. This is because while the second and fourth postulates together imply maximal entanglement between a portion of the the black hole interior and the late Hawking radiation, the first and third together imply that the late Hawking radiation must purify the early radiation as it is emitted. These constraints on the entanglement shared among the black hole, the early Hawking radiation, and the late Hawking radiation cannot be mutually satisfied without violating strong subadditivity of entanglement entropy. It would seem that taking all four postulates to be true leads to a contradiction, which must be resolved by requiring at least one of them to be violated in practice.

Several different resolutions to this puzzle have been proposed,[1] from those that modify quantum mechanics [245, 246], to those that allow a breakdown of no drama [18] or of unitarity [271], identify the early Hawking radiation with the black hole interior [174], modify the interior geometry [247, 250, 272], invoke quantum complexity theory [8, 253, 273], allow for black hole remnants [244], or take nonlocal approaches [248, 274].

By formulating black hole formation and evaporation as a process in Hilbert space in the context of Everettian quantum mechanics, we will argue that the four postulates above are made mutually consistent once we appreciate that the situations they refer to are not directly comparable. In particular, a prerequisite for both local effective field theory and no drama is the presence of a semiclassical background geometry. We will argue, as have several authors before us [272, 275–277], that in a fully quantum-gravitational treatment, an evaporating black hole is described

[1] This list is not meant to be exhaustive—for one listing see the comprehensive bibliography in Ref. [19].

not by a single semiclassical background but rather a superposition of many such geometries, each corresponding to different branches of the wave function.[2]

In short, while unitarity applies to the global wave function, the no-drama condition only applies on branches of the wave function. Therefore, the AMPS construction [18] does not lead to a paradox, as its components do not necessarily imply violation of monogamy of entanglement. Similar points have previously been made schematically [272,275–277,281–283], but in this work we will give a more precise articulation of this view. In doing so, we will also find that—under the reasonable assumption that each Hawking quantum has the opportunity to interact with the rest of the universe after it has been emitted—the production of a large number of decohered branches allows Hilbert space subfactors associated with the emitted radiation to have a large von Neumann entropy in the global wavefunction, even while they remain unentangled in every branch. This sort of entropic structure allows firewall-free individual branches while preserving the Page curve as a statement about the global wave function. Such a picture is heuristic at best, however, since the factorization of Hilbert space into early radiation, late radiation, and black hole degrees of freedom is highly branch-dependent.

According to the principle of black hole complementarity [23], we should not expect to be able to use local quantum field theory to simultaneously describe physics on both sides of a spacelike slice crossing an event horizon; what appear as local degrees of freedom inside a black hole will be distributed across the stretched horizon from the point of view of an external observer. Following our Hilbert-space perspective, we argue that this principle can be implemented in terms of how Hilbert space is factorized into subsystems and what basis of pointer states is associated with the resulting decomposition. The states that are robust with respect to environmental

[2]This conclusion could be viewed as a (mild and well-understood) violation of the second postulate above—there is not one local effective field theory for a single background but rather a different effective field theory on each semiclassical background. Because properties (such as the location) of the horizon can differ from branch to branch, our argument is reminiscent of state-dependent resolutions to the firewall paradox, e.g., Refs. [278–280]. We emphasize that this state-dependence arises naturally from the decoherence of the wave function and is not a violation of quantum mechanics, but rather a consequence of the fact that geometric properties differ from branch to branch.

monitoring from the point of view of an infalling observer will appear fragile to an outside observer. We exhibit an example decomposition of the relevant Hilbert spaces for each observer to show how this can work in practice.

Many puzzles about black hole evolution and evaporation certainly remain, such as whether the no-drama condition can be preserved at the level of the global wave function [284], how to reconstruct the black hole interior [246], and whether entanglement and wormholes are inextricably related [174]. Moreover, determining whether firewalls or smooth horizons with no drama are typical requires an analysis of the detailed branching structure of the global wave function for an evaporating black hole.

The rest of this paper is structured as follows. We begin in Sec. 11.2 by carefully formulating the process of black hole formation and evaporation so that we may properly discuss unitarity in a fully quantum-gravitational sense. Within this framework, we then investigate what it means to operationally probe entanglement between the black hole and exterior degrees of freedom in Sec. 11.3. We end with some brief concluding remarks in Sec. 11.4.

11.2 What is unitary and what is not

11.2.1 Setup

To examine unitarity for black hole formation and evaporation, let us set up the problem as a scattering experiment, employing the S-matrix ansatz [285] for asymptotically flat spacetime. Suppose that the initial state is a pure state of dilute matter that will collapse to form a black hole, specified on the past boundary of an asymptotically flat spacetime,

$$|\Psi_i\rangle \equiv |\Psi(i^- \cup \mathscr{I}^-)\rangle. \tag{11.1}$$

We define the initial state on the asymptotic past boundary so that it can be thought of as effectively some free-field-theoretic state without gravitational interactions.[3] If

[3] We will not consider the subtleties in the S-matrix formulation relating to infrared divergences; see [285, 286] and references therein.

quantum gravity is unitary, then this state unitarily evolves to another pure state; according to the S-matrix ansatz, the final pure state is given as a superposition of states each defined on the future boundary of an asymptotically flat spacetime,

$$|\Psi_f\rangle \equiv S|\Psi_i\rangle = \sum_j S_{i\to j}|\Psi_j(i^+ \cup \mathscr{I}^+)\rangle. \tag{11.2}$$

Although the asymptotically flat spacetimes, each corresponding to a branch j, are not identical, by definition each of them has the same boundary geometry (with $\mathscr{I}^+$ topology $S^{D-2} \times \mathbb{R}$). With an appropriate choice of coordinates, therefore, we can think of $|\Psi_f\rangle$ as a state that describes a superposition of definite field configurations on $i^+ \cup \mathscr{I}^+$. In general this time evolution is not described by a single Penrose diagram, since, in the bulk, the quantum-gravitational evolution of the wave function does not correspond to a single classical geometry.[4] Nevertheless, since the states at past and future null infinity are effectively noninteracting, we can identify all of these boundaries even in the absence of a well-defined bulk spacetime. A Penrose diagram for each individual process $S_{i\to j}$, if it exists, should look somewhat like the diagram sketched in Fig. 11.1: an asymptotically flat spacetime with some intermediate evaporating black hole geometry, the details of which we cannot resolve without an explicit understanding of quantum gravity.[5]

11.2.2 The Page curve: late-time entanglement structure

Consider factorizing the state $|\Psi_f\rangle$ as follows. Given a particular value of retarded time u on $\mathscr{I}^+$, with $u = +\infty$ corresponding to i^+, let us split the Hilbert space into the part with support to the past of u and the part with support to the future of u,

$$\mathcal{H}_{\mathscr{I}^+ \cup i^+} = \mathcal{H}_{<u} \otimes \mathcal{H}_{>u}. \tag{11.3}$$

The reduced state of the "early" Hawking radiation is then given by tracing over $\mathcal{H}_{>u}$,

$$\rho_{<u} = \mathrm{Tr}_{>u}\, |\Psi_f\rangle\langle\Psi_f|, \tag{11.4}$$

[4] The most general S-matrix setup would describe a wave function defined on some number of copies of $i^- \cup \mathscr{I}^-$ (only one for our choice of initial state $|\Psi_i\rangle$) that evolves to one defined on some number of copies of $i^+ \cup \mathscr{I}^+$, with no definitive spacetime structure in the interior.

[5] See also Fig. 5 of Ref. [277], which had previously advanced a similar view.

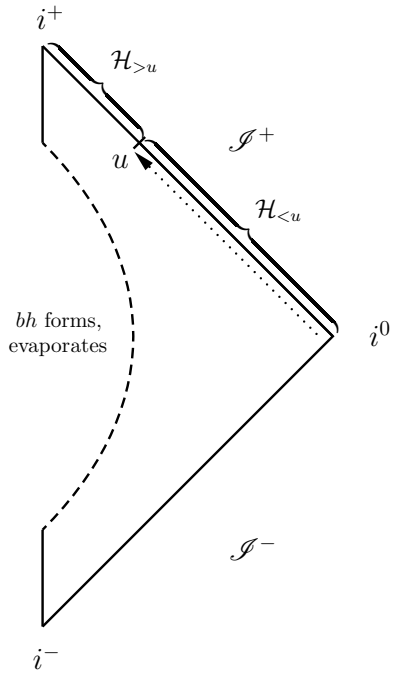

Figure 11.1: The Penrose diagram for the spacetime that corresponds to a classical branch of the global wave function that itself describes the unitary formation and evaporation of a black hole in asymptotically flat spacetime. On the asymptotic future boundary, we divide the global Hilbert space into two factors, $\mathcal{H}_{<u}$ and $\mathcal{H}_{>u}$, whose degrees of freedom lie to the past and future of the retarded time u, respectively. The direction of increasing u is indicated by the dotted arrow. The asymptotic future $i^+ \cup \mathscr{I}^+$ is identified across every classical branch of $|\Psi\rangle$ so that $\mathcal{H}_{<u}$ and $\mathcal{H}_{>u}$ are globally defined.

and the "Page curve" [40][6] is the plot of the von Neumann entropy of $\rho_{<u}$ as a function of u, which decreases to zero as u grows to cover all of $\mathscr{I}^+ \cup i^+$,

$$S(\rho_{<u})|_{u=+\infty} = 0. \tag{11.5}$$

That $S(\rho_{<u})$ vanishes when $u = +\infty$ is simply a consequence of unitary evolution, since the final state $|\Psi_f\rangle$ is correspondingly pure.[7] In other words, in the global wave function, the "late" Hawking radiation purifies the "early" radiation.[8]

[6] See Ref. [45] for further discussion of the Page curve.

[7] Maudlin [287] has recently emphasized that global unitary evolution is in principle consistent with information loss outside the black hole, since one can define disconnected Cauchy surfaces with respect to which the black hole interior persists as an effective "baby universe." We do not consider this possibility here, as it would violate Postulate 3, $S_{\rm bh} = S_{\rm BH}$. See also Ref. [269].

[8] We could have considered a spacetime with a timelike boundary, e.g., an asymptotically anti-de Sitter spacetime, but in that case defining the S-matrix proves difficult, for reasons discussed in for example Ref. [288].

We define the Page curve in terms of portions of the asymptotic future boundary because this definition does not rely on any particular choice of basis (for example, wave packets) for the Hawking radiation. We also remain agnostic about the exact shape of the Page curve resulting from this division of the final state into early and late radiation. Nevertheless, it is certainly true that $S(\rho_{<u})$ vanishes when $\rho_{<u}$ has support either nowhere or everywhere on the asymptotic future boundary.

11.2.3 Unitary evolution, branches, and decoherence

The modern black hole information problem arises when trying to interpret the entanglement structure at earlier times. Previously, we have only discussed the initial-state and late-time structure of the global wave function. However, because in this paper we are assuming that the (as yet unknown) theory of quantum gravity is a bona fide quantum-mechanical theory, we can also write down the wave function at intermediate times. Thus, the evolution of the state is, as usual, governed by the Schrödinger equation:[9]

$$\hat{H}|\Psi\rangle = i\frac{d}{d\lambda}|\Psi\rangle. \tag{11.6}$$

We emphasize that this equation genuinely implements time evolution; however, because λ need not have any relation to any coordinate or proper time in a geometric description,[10] we have chosen to use λ rather than t. Implementing our chosen boundary conditions, we must have $|\Psi(0)\rangle = |\Psi_i\rangle$ and $|\Psi(1)\rangle = |\Psi_f\rangle$. Because its evolution is governed by the Schrödinger equation, $|\Psi\rangle$ manifestly evolves unitarily.

A challenge in interpreting the state $|\Psi(\lambda)\rangle$ at intermediate values of λ lies in the fact that it does not describe a single black hole geometry. That is, a single geometry at one time (for example, $\lambda = 0$) must evolve to a state that describes an ensemble of many possible geometries at a later time. An observer or detector present in the

[9] In canonical quantum gravity, we could also take the point of view that the wave function should obey the Wheeler-DeWitt equation [289]. In this case, $\hat{H}$ is the Hamiltonian constraint, $\hat{H}|\Psi\rangle = 0$, and we need some additional information to implement time evolution as an emergent phenomenon. This approach is also proposed in Refs. [272, 277].

[10] In a holographic description, we could think of λ as the time coordinate of the boundary theory.

initial state would see different measurement outcomes depending on what geometry they were in at a later time.

For instance, while the expectation value of the black hole position and momentum remains fixed and constant in the global wave function, an observer who is monitoring the black hole would measure a drift in its position and momentum as it receives kicks from Hawking quanta that are emitted and interact with the surrounding environment, leading to decoherence. In Everettian language, the notion of a classical black hole geometry exists only on decohered branches of the global wave function. Therefore, in order to have an idea of a definite geometry throughout black hole formation and evaporation, it is necessary to specify what the decohered branches of the wave function are and what determines this branching structure.

This leads us to conjecture that the emergence of an ensemble of classical geometries from the unitary evolution of $|\Psi(\lambda)\rangle$ can be understood as a decoherence process, which determines a pointer basis for $|\Psi(\lambda)\rangle$ whose elements describe decohered geometries and configurations of matter. The lesson of the decoherence program [290–294] is that branching of the wave function is set by the interaction dynamics between a particular subsystem and the environment monitoring this subsystem. In order to determine the branching structure, we need to decompose the Hilbert space into "system" and "environment" degrees of freedom. For our purposes, we suppose that there exists a set of degrees of freedom in the total Hilbert space that can serve as an environment that, minimally, yields a definite geometry when traced over. For example, one could conjecture this environment to be comprised of some inherently quantum gravitational degrees of freedom.[11] Alternatively, recent studies suggest that the modes of soft gravitons and other soft massless gauge bosons may constitute such an environment [296–301]. Regardless, we conjecture that the global Hilbert space may be written as

$$\mathcal{H} = \mathcal{H}_{\text{eff}} \otimes \mathcal{H}_{\text{env}}, \tag{11.7}$$

[11] For some discussion of this kind of UV/IR factorization, see Refs. [21, 172, 295].

so that the global state takes the form of a sum over branches,

$$|\Psi(\lambda)\rangle = \sum_b \alpha_b(\lambda)|\Psi_b\rangle, \tag{11.8}$$

where each branch $|\Psi_b\rangle$ decomposes as

$$|\Psi_b\rangle = |\psi_b\rangle_{\text{eff}} \otimes |\varepsilon_b\rangle_{\text{env}} \,. \tag{11.9}$$

Then $|\psi_b\rangle_{\text{eff}} \in \mathcal{H}_{\text{eff}}$ is the part of the state that we may think of as describing a semiclassical geometry and the states of quantum fields in the theory on top of this geometry, while $|\varepsilon_b\rangle_{\text{env}}$ is a state of the environmental degrees of freedom that are responsible for decohering the state to a semiclassical geometry. How the global wave function branches depends on how the $|\varepsilon_b\rangle_{\text{env}}$ are determined.

Depending on the superselection rules of quantum gravity, we might only need to consider, e.g., branches b that correspond to asymptotically flat geometries, geometries with identical topologies to the initial state, etc. We can either implement these rules by working in a smaller Hilbert space than the full Hilbert space of quantum gravity or by imposing that $\alpha_b(\lambda) = 0$ for all branches b corresponding to geometries that do not obey these superselection rules.[12]

Trivially, the Hilbert space $\mathcal{H}_{\text{eff}}$ admits a direct sum structure [113, 303],

$$\mathcal{H}_{\text{eff}} = \bigoplus_b \text{span}\{|\psi_b\rangle_{\text{eff}}\}. \tag{11.10}$$

However, we expect that it should be possible to group sets of states together that have the same background geometry to form subspaces

$$\mathcal{H}^{\mathcal{B}}_{\text{eff}} \equiv \text{span}\left\{|\psi_b\rangle_{\text{eff}} \mid b \in \mathcal{B}\right\}, \tag{11.11}$$

where any given $\mathcal{B}$ contains the labels of a set of branches that all correspond to the same background geometry (to within some precision that specifies a coarse-graining). As such, we envision each $\mathcal{H}^{\mathcal{B}}_{\text{eff}}$ as being the Hilbert space of fields coupled to the background geometry of the branches in $\mathcal{B}$. Each $\mathcal{H}^{\mathcal{B}}_{\text{eff}}$ can of course be further

[12] For example, thought experiments in AdS/CFT suggest the existence of topological superselection rules for holographic wormhole geometries [302].

decomposed, for example, as a tensor product over the Hilbert spaces of the different species of fields contained in the effective theory. Because field theories are defined on fixed spacetime backgrounds, a tensor product in each individual $\mathcal{H}^{\mathcal{B}}_{\text{eff}}$ does not necessarily extend to a tensor product on the entire semiclassical Hilbert space $\mathcal{H}_{\text{eff}}$. Generally we can therefore write

$$\mathcal{H}_{\text{eff}} = \bigoplus_{\mathcal{B}} \mathcal{H}^{\mathcal{B}}_{\text{eff}} = \bigoplus_{\mathcal{B}} \left(\bigotimes_{i} \mathcal{H}_i^{(\mathcal{B})} \right), \tag{11.12}$$

where the $\mathcal{H}_i^{(\mathcal{B})}$ represent factors defined on the Hilbert space of the specific background geometry $\mathcal{B}$. In particular, notions such as "modes of outgoing Hawking radiation near the horizon" are only well-defined on specific branches, not on the global wave function.

During the process of decoherence itself, the action of the Hamiltonian entangles system and environment states and the entropy of the system density operator $\rho_{\text{eff}} = \text{Tr}_{\text{env}} |\Psi\rangle\langle\Psi|$ increases. After decoherence, ρ_{eff} will be diagonal with respect to a basis of "pointer states" for $\mathcal{H}_{\text{eff}}$, each pointer state defining a different branch of the wave function. For us, the pointer states are the $\{|\psi_b\rangle\}$, representing quantum fields on a definite semiclassical background. Once this occurs, branches interact minimally with each other (they decohere), so that the time evolution of a superposition of branches is approximately the same as evolving each branch individually. In particular, the branches retain their product-state structure (11.9) under the action of the Hamiltonian implementing time evolution.

Returning to Eq. (11.8), we see that at each time λ there are a number of decohered branches, describing a superposition of the geometries corresponding to those $|\psi_b\rangle_{\text{eff}}$ with $\alpha_b(\lambda) \neq 0$. As λ increases, so does the number of decohered branches, i.e., the size of the set $\{|\Psi_b\rangle | \alpha_b(\lambda) \neq 0\}$. It seems natural to relate this repeated branching to the production of entropy. As a result, the increase in the number of decohered branches is important for the interpretation of the Hawking entropy formula and the Page curve, as we discuss in Sec. 11.2.5.

11.2.4 Entanglement structure at intermediate times

The basic reason why the consideration of branching structure is relevant to the black hole information puzzle is that, while evolution of the global wave function is unitary, evolution via conditioning on a specific background geometry (i.e., projection onto individual branches of the wave function) is not. A particular sequence of classical states is produced by repeated non-unitary projection of the wave function onto states of definite background geometry; we refer colloquially to wave function "collapse" during the measurement process (which for us is simply decoherence and branching). In particular, the Page curve, which we have seen above is a consequence of unitarity, only needs to hold for the global wave function.

Our main observation is that arguments for the modern information puzzle—and in particular Postulates 2 and 4 above—only apply at the level of the $|\psi_b\rangle_{\text{eff}}$ parts of the classical branches [275, 276]. Again, evolution of the global wave function is unitary, but evolution at the level of individual classical geometries is not. In Sec. 11.3, we will discuss what it means to operationally probe the information puzzle in the context of this observation. In essence, at intermediate times, it is not clear how to calculate the Page curve as we have formulated it in Sec. 11.2.2 because the specification of what degrees of freedom constitute "early" radiation is a branch-dependent notion.

One possibility could be to make a branch-dependent local tensor product decomposition of the type suggested by AMPS, where $\mathcal{H}_{\text{eff}}$ is taken to be $A^{(b)} \otimes B^{(b)} \otimes R^{(b)} \otimes C^{(b)}$, where $R^{(b)}$ denotes degrees of freedom that correspond to early radiation, $A^{(b)}$ corresponds to the black hole degrees of freedom, $B^{(b)}$ is the late radiation, and $C^{(b)}$ (for "complement") is everything else. However, such a decomposition is problematic for a number of reasons. Generically, it will not be the case that $R^{(b)}$, $B^{(b)}$, $A^{(b)}$ are the same factors on every classical branch $|\Psi_b\rangle$. For instance, we can identify a space $A^{(b)}$ of black hole degrees of freedom on branches where a black hole exists and it is likely that this space may be consistently identified across all branches $b \in \mathcal{B}$ with the same background geometry. But we cannot speak of anything like a global space of black hole degrees of freedom in the Hilbert space of all semiclassical states. Moreover,

even within branches that describe the same background geometry, Hilbert space subfactors that consist of quanta of field excitations, such as $R^{(b)}$, will vary from branch to branch. (For example, a branch with one decohered graviton of energy E and a branch with two decohered gravitons of energies $E_1 + E_2 = E$ are distinct.)

Furthermore, a decomposition such as $A^{(b)} \otimes B^{(b)} \otimes R^{(b)} \otimes C^{(b)}$ becomes a highly "observer-dependent" refinement, in the sense that the Hilbert space factors are neither dictated by the theory itself nor have a fixed spacetime interpretation. For example, according to the description of an observer outside of a black hole on a branch b, states in $A^{(b)}$ describe states of the black hole's stretched horizon. For such an observer, the interior of the black hole is not a geometric place, which runs counter (but complementary) to the description of an infalling observer if the horizon is transparent. We will return to the question of complementarity and further decomposition of the classical branches in $\mathcal{H}_{\text{eff}}$ in Sec. 11.3, but for now we stress that none of our arguments assume that the exterior observer can assign any classical geometric interpretation to the black hole interior.

One of our main conclusions about black hole evolution in the global wave function is that unitary evolution and the no-drama condition are compatible in principle, even without violating monogamy of entanglement. Unitarity is a global concept; it applies only to the global wave function and not to individual semiclassical branches of particular geometry. On the other hand, the requirement of no drama is a statement about individual decohered branches describing such semiclassical geometries; in particular, it is a requirement that the state of the quantum fields near the black hole take a particular structure (corresponding to the vacuum) *on the branch*, i.e., within $|\psi_b\rangle_{\text{eff}}$, for most of the branches.

Let us consider a toy model to illustrate that what parties appear entangled on decohered branches of a wave function can be very different from the structure of entanglement entropies in the global wave function. Consider, for example, four qubits labeled A, B, C, and D in the state

$$|\Psi\rangle_{ABCD} = \tfrac{1}{\sqrt{2}} \left(|00\rangle_{AD} + |11\rangle_{AD}\right) \otimes \tfrac{1}{\sqrt{2}} \left(|00\rangle_{BC} + |11\rangle_{BC}\right). \tag{11.13}$$

In this state and tensor decomposition, the pairs AD and BC are unentangled, while A and D, as well as B and C, are entangled. However, suppose that we treat CD as an environment and posit a Hamiltonian with an interaction term between AB and CD of the form

$$H_{\text{int}} = \sum_{\pm} \mathcal{O}_{AB}^{(1),\pm} \otimes |\phi^{\pm}\rangle\langle\phi^{\pm}|_{CD} + \mathcal{O}_{AB}^{(2),\pm} \otimes |\psi^{\pm}\rangle\langle\psi^{\pm}|_{CD}, \tag{11.14}$$

where $|\phi^{\pm}\rangle = \frac{1}{\sqrt{2}}(|00\rangle \pm |11\rangle)$ and $|\psi^{\pm}\rangle = \frac{1}{\sqrt{2}}(|01\rangle \pm |10\rangle)$ denote Bell states. Then it follows that the branching structure of $|\Psi\rangle_{ABCD}$ (whose environmental CD parts commute with the interaction Hamiltonian) looks like

$$\begin{aligned} |\Psi\rangle_{ABCD} = \frac{1}{2}\big(&|\phi^+\rangle_{AB} \otimes |\phi^+\rangle_{CD} + |\phi^-\rangle_{AB} \otimes |\phi^-\rangle_{CD} \\ &+ |\psi^+\rangle_{AB} \otimes |\psi^+\rangle_{CD} - |\psi^-\rangle_{AB} \otimes |\psi^-\rangle_{CD}\big)\,. \end{aligned} \tag{11.15}$$

In other words, while AD and BC are unentangled in the global wave function, in the sense that $S(AD) = 0$, the "system" subfactors A and B are entangled on every branch. Tracing out CD to obtain a reduced density matrix for AB would reveal a set of distinct branches, all of which exhibit maximal entanglement between A and B.

More generally, the production of a large number of orthogonal branches through decoherence can lead to large von Neumann entropies for subsystems in the global wave function. Heuristically, this is what we expect to happen for AMPS-like tensor product factors; $A^{(b)}$ and $B^{(b)}$ must be highly entangled for drama-free branches, yet decoherence can produce large von Neumann entropies for the collections of $A^{(b)}B^{(b)}$ and $R^{(b)}$ on the branches, due to classical uncertainty. This is only a heuristic picture, since there is no consistent identification of AMPS-like tensor product factors across all branches that may be used to compute the Page curve at intermediate times. Nevertheless, it is interesting to assess just how much entropy is produced by branching, e.g., within a given sector $\mathcal{B}$, which we now discuss.

11.2.5 Branch counting

Consider a simple idealization, according to which $AB \subset U^{\dagger}(\lambda;1)[\mathcal{H}_{>u}]$ and $R \subset U^{\dagger}(\lambda;1)[\mathcal{H}_{<u}]$ actually are consistently identified as the same factors across all branches

(even though, as discussed above, that's not precisely the case in our scenario). Here $U(\lambda_1;\lambda_2)$ is the unitary evolution operator that maps a state at parameter value λ_1 to the state at parameter value $\lambda_2 > \lambda_1$. In other words, here we explicitly hypothesize that the Hilbert-space decomposition $(A^{(b)} \otimes B^{(b)} \otimes R^{(b)} \otimes C^{(b)})$ holds globally across all branches and we explore the resulting consequences.[13]

Tracing over $\mathcal{H}_{\text{env}}$ and C in the global wave function, the reduced state on ABR can take the form

$$\text{Tr}_{\text{env},C}\,|\Psi(\lambda)\rangle\langle\Psi(\lambda)| = \sum_b p_b(\lambda)\,\rho_b^{AB} \otimes \rho_b^R, \tag{11.16}$$

where, on each branch, AB and R are unentangled (even though they are correlated globally). Such entanglement structure is required in order to avoid, for example, a firewall arising from broken entanglement across the AB subsystems between the black hole and outgoing late radiation modes. That is, on each branch defining a classical spacetime geometry, we let the quantum fields take the vacuum configuration at the horizon, as required by Postulate 4.

Even though AB and R are unentangled on every branch, there is still nonzero von Neumann entropy for AB and R globally. Consider the reduced state on AB alone,

$$\rho^{AB}(\lambda) = \sum_b p_b(\lambda)\rho_b^{AB}. \tag{11.17}$$

The Holevo information [304, 305] of ρ^{AB} is given by

$$\chi(\rho^{AB}) = S(\rho^{AB}) - \sum_b p_b\, S(\rho_b^{AB}) \tag{11.18}$$

and is an upper bound on the accessible information of ρ^{AB} and its corresponding ensemble. More importantly for our purposes, it is bounded by the Shannon entropy, $-\sum_b p_b \log p_b$, with saturation occurring when each ρ_b^{AB} has orthogonal support [304]. Moreover, $S(\rho^{AB})$ can be bounded from below by using the concavity of

[13] Alternatively, we can think of the mental exercise discussed here as taking place within a collection of branches $\mathcal{B}$ that initially have the same geometry: We first project onto a collection of branches $|\Psi_b\rangle$, $b \in \mathcal{B}$, on which we make this decomposition of $\mathcal{H}_{\text{eff}}$ into $ABRC$ and we then study further evolution of entanglement.

entanglement entropy. Putting these bounds together, we have

$$\sum_b p_b\, S(\rho_b^{AB}) \leq S(\rho^{AB}) \leq \sum_b p_b\, S(\rho_b^{AB}) - \sum_b p_b \log p_b\,. \tag{11.19}$$

In particular, $S(\rho^{AB})$ can in fact be quite large. For example, in the case where each ρ_b^{AB} has orthogonal support, then $S(\rho^{AB}) \approx \log N$ if each $p_b \approx 1/N$, where N is the number of branches (i.e., the sum over b runs from 1 to N). An old black hole of mass M will have been emitting Hawking quanta of typical energy $\lesssim 1/M$, so greater than $O(M/(1/M)) = O(M^2)$ emissions will have occurred in the black hole's past. If each emission branches the global wave function by a constant factor, then the scaling of N goes as e^{M^2}. In order to specify a branch, we must choose not only the mass and momentum of the black hole itself, but the entire exterior spacetime geometry, which, via back reaction, depends on the distribution of all the Hawking radiation between the black hole and $\mathscr{I}^+$. It is therefore plausible that each Hawking emission indeed branches the global wave function, as long as the emitted quantum becomes entangled with the environment.[14] Had we only considered the macroscopic properties of the black hole itself, the number of branches would be much smaller [19]. However, as long as there is anything else in the universe aside from the black hole for a Hawking quantum to interact with on its way to infinity, it is reasonable to treat the back reaction of the Hawking quantum on the spacetime as decohering the wave function into states of definite geometry. Even with no exterior matter outside the black hole, it is conceivable that the gravitational interaction of Hawking quanta is itself enough to decohere the geometry; an exploration of this question and its possible connection to recent work on soft gravitons and their associated symmetries [299] lies beyond the scope of the present work. Note that, in our setup, $S(\rho^{AB})$ scales in the same way as the Bekenstein-Hawking entropy of the black hole, $S_{\rm BH} \sim M^2$, so it may be possible to recover Postulate 3, $S_{\rm bh} = S_{\rm BH}$, via the branching structure alone,[15] but the details of the branch counting also lie

[14] If Hawking quanta never become entangled with something that could be labeled "an environment," branching would not occur. In that case, however, there is no sensible way to assign a semiclassical geometry to the state and it is not appropriate to speak of a black hole, much less a firewall.

[15] We note that a similar argument was made in the context of the fuzzball program in Ref. [306].

beyond the scope of the present work. Such an analysis of the branching structure would be necessary in order to guarantee that no-drama states are indeed generic for a randomly-selected black hole horizon in the global wave function; here we merely want to emphasize that such states are plausible.

11.3 Operational tests of the information puzzle

We now turn to the question of how a pair of observers would practically implement the AMPS thought experiment [18] to probe the state of the black hole inside and outside the event horizon. Our main concern is to understand this implementation in the context of unitarity of the global wave function. In particular, we will argue that the branching structure of the global wave function is such that the state vectors that are robust under Hamiltonian evolution—the *pointer states* into which the global wave function branches—are very different inside and outside the horizon. Specifically, the pointer bases corresponding to measurements made by an interior, infalling observer and an external, static observer are related to each other via non-trivial transformations that manifest the complexity of black hole scrambling. This means that it is impossible for both the infalling observer behind the horizon and the external observer to exist on the same semiclassical branch of the wave function at their level of coarse-graining.[16]

Let us suppose that, as part of the initial asymptotic data, we specify that there are two detectors at i^-, $D_{\rm inf}$ and $D_{\rm st}$, corresponding to infalling and stationary observers and that each begins in some ready state d_0. Assume that the detectors are local, can be switched on and off, and are identical in operation. We can decompose the effective Hilbert space as

$$\mathcal{H}_{\rm eff} = \widetilde{\mathcal{H}}_{\rm eff} \otimes D_{\rm inf} \otimes D_{\rm st}. \tag{11.20}$$

Specifically, we isolate the finite-dimensional Hilbert spaces $D_{\rm inf}$ and $D_{\rm st}$ that represent the detectors' internal degrees of freedom that ultimately couple to some local system to realize measurement. We suppose that all of the detectors' other degrees

[16] See Ref. [281] for a discussion of related ideas.

of freedom, such as kinematic degrees of freedom like position and momentum, are a part of $\widetilde{\mathcal{H}}_{\rm eff}$. At intermediate parameter values λ with the detectors switched off, we therefore write the global wave function as

$$|\Psi(\lambda)\rangle = \sum_b \alpha_b(\lambda)|\tilde{\psi}_b\rangle_{\rm eff} \otimes |d_0\rangle_{D_{\rm inf}} \otimes |d_0\rangle_{D_{\rm st}} \otimes |\varepsilon_b\rangle_{\rm env}\,. \tag{11.21}$$

Our aim is to consider a situation in which one detector, $D_{\rm st}$, remains stationary outside of a black hole and the other, $D_{\rm inf}$, falls into the same black hole and to compare the measurements reported by the two detectors. We therefore begin by projecting onto a branch of the wave function corresponding to a single spacetime so that the two detectors agree on the background geometry. The object of interest is thus a particular branch b_* of the form $|\tilde{\psi}_{b_\star}\rangle_{\rm eff} \otimes |d_0\rangle_{D_{\rm inf}} \otimes |d_0\rangle_{D_{\rm st}}$. For convenience, we have temporarily dropped the $\mathcal{H}_{\rm env}$ factor because it plays no role once we have projected onto a branch (keeping in mind that $\mathcal{H}_{\rm env}$ is necessary for further evolution of the initial branch to be unitary).

Equipped with a notion of background spacetime, we can now attempt to interpret $\mathcal{H}_{\rm eff}$ in terms of spacetime regions and in the context of measurements performed by the detectors on the branch $b_\star$. Suppose that, on the branch in question, $D_{\rm inf}$ falls into the black hole while $D_{\rm st}$ remains outside. Further suppose that, on this branch, at some moment, both detectors switch on and become entangled with the local degrees of freedom that they probe. Let us define factors of $\mathcal{H}_{\rm eff}$ on a Cauchy surface chosen such that its intersection with the infalling detector's worldline occurs inside the black hole.

Consider first the following decomposition of $\mathcal{H}_{\rm eff}$, appropriate from the point of view of the stationary detector:

$$\begin{aligned}\mathcal{H}_{\rm eff} &= A \otimes S \otimes E \otimes D_{\rm st} \\ &\equiv S \otimes D_{\rm st} \otimes \mathcal{E}.\end{aligned} \tag{11.22}$$

Here,

- A is the black hole Hilbert space, which, in the spirit of complementarity, we suppose represents states of the stretched horizon,
- $D_{\rm st}$ is the Hilbert space of the stationary detector,
- S is the collection of local degrees of freedom that constitute the system that the stationary detector measures, and
- E are any remaining exterior degrees of freedom. Altogether, $\mathcal{E} \equiv A \otimes E$ is the environment for the stationary detector.

Similarly, we can also decompose $\mathcal{H}_{\rm eff}$ in a way that is appropriate for an infalling description:

$$\begin{aligned} \mathcal{H}_{\rm eff} &= D_{\rm inf} \otimes T \otimes F \otimes G \\ &\equiv D_{\rm inf} \otimes T \otimes \mathcal{F}. \end{aligned} \tag{11.23}$$

Here,

- $D_{\rm inf}$ is the Hilbert space of the infalling detector,
- T is the system that the interior detector measures, and
- F and G are other degrees of freedom inside and outside the black hole, respectively. Altogether, $\mathcal{F} \equiv F \otimes G$ is the environment for the stationary detector.

How the various Hilbert space decompositions overlap is illustrated in Fig. 11.2.

One of the key results of black hole complementarity is that horizon dynamics, as seen by a stationary exterior observer, appear to be approximately typical with respect to the Haar measure on sufficiently long timescales. This is discussed in, for example, Refs. [253, 257].[17] In the remainder of this section, we will find that black hole scrambling has important implications for the structure of the global wave function in terms of the pointer bases of interior and exterior observers.

[17] Ref. [133] discusses classical black hole scrambling in the context of the membrane paradigm.

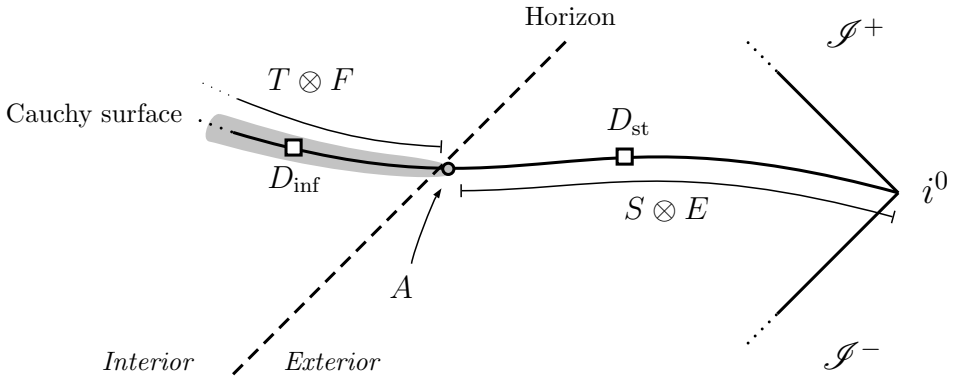

Figure 11.2: Diagrammatic representation of the decompositions of $\mathcal{H}_{\text{eff}}$ in Eqs. (11.22) and (11.23) on a Penrose diagram representing some particular semi-classical branch $b_\star$. The detectors and their associated internal Hilbert spaces, D_{st} and D_{inf}, are denoted by the white boxes. The location of the stretched horizon and its associated Hilbert space, A, is denoted by the gray circle. According to black hole complementarity, we suppose that $A = D_{\text{inf}} \otimes T \otimes F$ are identified as the same Hilbert space. This is indicated by the shading of the part of the Cauchy surface in the black hole interior. The interior Hilbert space factors and the interior geometry are only resolved by observers who cross the black hole's horizon. From the point of view of an exterior observer, these degrees of freedom are precisely the degrees of freedom of the stretched horizon. Also note that, according to Eqs. (11.22) and (11.23), $G = D_{\text{st}} \otimes S \otimes E$.

Let us first develop some intuition for what to expect. Consider an infalling object crossing the stretched horizon as seen by either an observer falling along with the object or a stationary observer at some fixed position outside of the black hole. While for the infalling observer the object will seem to pass through the horizon without any apparent effect, for the external observer the object will appear to scramble across and thermalize with the stretched horizon. If the infalling object is a classical object—which in particular means that it is robust against decoherence due to monitoring by its environment in the infalling frame—this picture suggests that the object is explicitly not robust against decoherence due to environmental interactions *in the frame of the external observer*, in which it is seen to quantum-mechanically scramble and delocalize across the entirety of the stretched horizon. As the states that are robust against decoherence are by definition the pointer states, this highlights the fact that the pointer states in the infalling frame, when viewed in the frame of the static external observer, appear to be totally scrambled and delocal-

ized. In the context of black holes, this feature has traditionally been implemented by a unitary 2-design [42], which up to its second moment is indistinguishable from a Haar-typical unitary.

Consider decomposing the particular state $|\psi_{b_\star}\rangle_{\text{eff}}$ according to the two branching structures implied by the two detectors and their decohering dynamics:

$$\begin{aligned} |\psi_{b_\star}\rangle_{\text{eff}} &= \sum_i c'_i |s_i\rangle_S |d_i\rangle_{D_{\text{st}}} |\alpha_i\rangle_{\mathcal{E}} \qquad \text{(stationary)} \\ |\psi_{b_\star}\rangle_{\text{eff}} &= \sum_j c''_j |s_j\rangle_T |d_j\rangle_{D_{\text{inf}}} |\beta_j\rangle_{\mathcal{F}} \qquad \text{(infalling)}\,, \end{aligned} \tag{11.24}$$

where "stationary" and "infalling" remind us whether we are expressing the state in the basis of the detector stationary outside or infalling inside the black hole. Let us focus on the branching structure from the stationary point of view and start writing $\mathcal{E}$ in terms of Hilbert space subfactors as $|\alpha_i\rangle_{\mathcal{E}} = \sum_{kl} \mu^i_{kl} |\phi_k\rangle_A |e_l\rangle_E$. A state $|\phi_k\rangle_A$ describing the stretched horizon can be decomposed into its constituent factors in D_{inf}, T, and F,

$$|\phi_k\rangle_A = \sum_{abc} U^k_{abc} |s_a\rangle_T |d_b\rangle_{D_{\text{inf}}} |f_c\rangle_F, \tag{11.25}$$

where U^k_{abc} implements a unitary change of basis from abc to k. Black hole scrambling implies that this unitary is, to a good approximation, Haar-typical for a generic choice of basis for A and its constituent Hilbert-space factors. We therefore have

$$|\psi_{b_\star}\rangle_{\text{eff}} = \sum_i c'_i |s_i\rangle_S |d_i\rangle_{D_{\text{st}}} \sum_{kl} \mu^i_{kl} \left(\sum_{abc} U^k_{abc} |s_a\rangle_T |d_b\rangle_{D_{\text{inf}}} |f_c\rangle_F \right) |e_l\rangle_E. \tag{11.26}$$

Rearranging the sums, we have

$$\begin{aligned} |\psi_{b_\star}\rangle_{\text{eff}} &= \sum_{ab} |s_a\rangle_T |d_b\rangle_{D_{\text{inf}}} \sum_c \left[\sum_i c'_i \left(\sum_l \left[\sum_k \mu^i_{kl} U^k_{abc} \right] |e_l\rangle_E \right) |s_i\rangle_S |d_i\rangle_{D_{\text{st}}} \right] |f_c\rangle_F \\ &= \sum_a c''_a |s_a\rangle_T |d_a\rangle_{D_{\text{inf}}} |\beta_a\rangle_{\mathcal{F}}. \end{aligned} \tag{11.27}$$

To recap: In Eq. (11.26), we wrote each $|\alpha_i\rangle_{\mathcal{E}}$ in an orthonormal basis for the horizon (A) and E, taking each horizon state and expanding it in the pointer state basis for T and D_{inf}, along with some arbitrary basis for F. We can also express $|\psi_{b_\star}\rangle_{\text{eff}}$ in the pointer basis of the infalling detector, writing it with the branching structure as

given in the second line of Eq. (11.27). Hence, if both the infalling and stationary detector have decohered, it must be that $U^k_{abc} = 0$ if $a \neq b$ so that

$$\begin{aligned}|\psi_{b_\star}\rangle_{\text{eff}} &= \sum_a |s_a\rangle_T |d_a\rangle_{D_{\text{inf}}} \sum_c \left[\sum_i c'_i \left(\sum_l \left[\sum_k \mu^i_{kl} U^k_{aac}\right] |e_l\rangle_E\right) |s_i\rangle_S |d_i\rangle_{D_{\text{st}}}\right] |f_c\rangle_F \\ &\equiv \sum_a |s_a\rangle_T |d_a\rangle_{D_{\text{inf}}} \underbrace{\sum_c \left[\sum_i c'_i |\tilde{e}^i_{ac}\rangle_E |s_i\rangle_S |d_i\rangle_{D_{\text{st}}}\right] |f_c\rangle_F}_{c''_a |\beta_a\rangle_{\mathcal{F}}} .\end{aligned} \tag{11.28}$$

That the horizon scrambles means that the components U^k_{aac} are approximately typical with respect to the Haar measure.

Were we to find that $\sum_k \mu^i_{kl} U^k_{aac} \propto \delta^i_a$, then the sum in Eq. (11.28) would collapse to a single term:

$$|\psi_{b_\star}\rangle_{\text{eff}} = \sum_a |s_a\rangle_T |d_a\rangle_{D_{\text{inf}}} c'_a |s_a\rangle_S |d_a\rangle_{D_{\text{st}}} \left(\sum_c |\tilde{e}^a_{ac}\rangle_E |f_c\rangle_F\right) . \tag{11.29}$$

Such a situation would be pathological because it would mean that pointer states of the black hole exterior would correlate perfectly with states of the black hole interior, which would mean that they would be stable under the action of their joint environment, i.e., classical and long-lived. This would seem to contradict what is believed about black hole fast scrambling. Moreover, such a conspiracy between the matrices μ and U is implausible since U is Haar-typical and furthermore dependent on the detector that we choose. To see this, note that U describes how the state of the stretched horizon decomposes in the infalling detector's pointer basis, while μ is independent of the detector properties, simply describing the joint state of the stretched horizon and exterior environment, and has no reason to be correlated with the Haar-typical properties of U.

Thus, we have shown that the pointer bases for the interior and exterior observer are not compatible. Specifically, Eq. (11.28) shows that the environment states $|\beta_a\rangle_{\mathcal{F}}$ for the infalling detector are given by nontrivial transformations (under $\mu^i_{kl} U^k_{aac}$) of the joint state of the exterior system, detector, and environment, along with the interior environment. Similarly, the environment states associated with the pointer basis for

the exterior detector are given by nontrivial transformations of the joint state of the interior system, detector, and environment, along with the exterior environment.

What this means physically is that it is not possible to isolate a single branch of the wave function, via a natural dynamical decoherence process, that corresponds to a pointer state for the interior and exterior detector simultaneously. This property of the global wave function reconciles the complementary points of view of infalling and stationary observers, without requiring the existence of a firewall to preserve unitarity. Black hole complementarity is therefore implemented in Hilbert space in terms of the relationship between pointer states as defined by different observers across a horizon.

11.4 Conclusions

The information paradox, as sharpened by AMPS, seemingly necessitated modifying a cherished pillar of modern physics in effective field theory, relativity, or unitarity. In this work, we argued using decoherence and pointer bases that such a dramatic conclusion is not directly implied by the ingredients of the AMPS discussion. In particular, different components of the AMPS argument apply either globally or on individual branches of the wave function: unitarity applies to the global wave function, while the absence of drama at the horizon is a statement about individual semiclassical branches. They can therefore, as far as we can tell, be satisfied simultaneously without violating unitarity, monogamy of entanglement, or any other principles of quantum mechanics.

Since the existence of firewalls would stand in gross violation of our classical intuition, we should judge them to be unlikely unless their absence would require violating an even-more-cherished belief, which we have argued it does not. Given our best current understanding of quantum mechanics and black hole thermodynamics, there is no reason to insist that an observer falling into a black hole sees anything other than a reason to regret their decision.

Acknowledgements

We thank Ahmed Almeiri, Raphael Bousso, William Donnelly, Masahiro Hotta, Cindy Keeler, Yasunori Nomura, Don N. Page, Guillaume Verdon, and Koji Yamaguchi for helpful discussions. This work was supported by the U.S. Department of Energy, Office of Science, Office of High Energy Physics, under Award Number DE-SC0011632. N.B. was supported by the National Science Foundation, under grant number 82248-13067-44-PHPXH. A.C.-D. was supported by a Beatrice and Sai-Wai Fu Graduate Fellowship in Physics and the Gordon and Betty Moore Foundation through Grant 776 to the Caltech Moore Center for Theoretical Cosmology and Physics. J.P. was supported in part by the Simons Foundation and in part by the Natural Sciences and Engineering Research Council of Canada. G.N.R. was supported by the Miller Institute for Basic Research in Science at the University of California, Berkeley.

BIBLIOGRAPHY

[1] N. Bao and A. Chatwin-Davies, "Puzzles and pitfalls involving Haar-typicality in holography," arXiv:1708.08561.

[2] N. Bao and A. Chatwin-Davies, "The complexity of identifying Ryu-Takayanagi surfaces in AdS3/CFT2," *JHEP* **11** (2016) 034, arXiv:1609.01727.

[3] N. Bao, C. Cao, S. M. Carroll, A. Chatwin-Davies, N. Hunter-Jones, J. Pollack, and G. N. Remmen, "Consistency conditions for an AdS multiscale entanglement renormalization ansatz correspondence," *Phys. Rev. D* **91** (2015) 125036, arXiv:1504.06632.

[4] N. Bao, C. Cao, S. M. Carroll, and A. Chatwin-Davies, "De Sitter space as a tensor network: Cosmic no-hair, complementarity, and complexity," *Phys. Rev. D* **96** (2017) 123536, arXiv:1709.03513.

[5] S. M. Carroll and A. Chatwin-Davies, "Cosmic equilibration: A holographic no-hair theorem from the generalized second law," *Phys. Rev. D* **97** (2018) 046012, arXiv:1703.09241.

[6] A. Chatwin-Davies, A. Kempf, and R. T. W. Martin, "Natural covariant Planck scale cutoffs and the cosmic microwave background spectrum," *Phys. Rev. Lett.* **119** (2017) 031301, arXiv:1612.06445.

[7] A. Chatwin-Davies, A. S. Jermyn, and S. M. Carroll, "How to recover a qubit that has fallen into a black hole," *Phys. Rev. Lett.* **115** (2015) 261302, arXiv:1507.03592.

[8] N. Bao, A. Bouland, A. Chatwin-Davies, J. Pollack, and H. Yuen, "Rescuing complementarity with little drama," *JHEP* **12** (2016) 026, arXiv:1607.05141.

[9] N. Bao, S. M. Carroll, A. Chatwin-Davies, J. Pollack, and G. N. Remmen, "Branches of the black hole wave function need not contain firewalls," arXiv:1712.04955.

[10] D. E. Bruschi, T. Ralph, I. Fuentes, T. Jennewein, and M. Razavi, "Spacetime effects on satellite-based quantum communications," *Phys. Rev. D* **90** (2014) 045041, arXiv:1309.3088.

[11] P. Stamp, "Correlated worldline theory of quantum gravity: low-energy consequences & table-top tests." Talk given at the Caltech IQI Seminar, November, 2015.

[12] Q. Lin, J. Rosenberg, D. Chang, R. Camacho, M. Eichenfield, K. J. Vahala, and O. Painter, "Coherent mixing of mechanical excitations in nano-optomechanical structures," *Nat. Photonics* **4** (2010) 236–242, arXiv:0908.1128.

[13] J. D. Bekenstein, "Generalized second law of thermodynamics in black hole physics," *Phys. Rev. D* **9** (1974) 3292–3300.

[14] R. Bousso, "A covariant entropy conjecture," *JHEP* **07** (1999) 004, arXiv:hep-th/9905177.

[15] J. M. Maldacena, "The large N limit of superconformal field theories and supergravity," *Int. J. Theor. Phys.* **38** (1999) 1113–1133, arXiv:hep-th/9711200. [*Adv. Theor. Math. Phys.* **2** (1998) 231].

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

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

[18] A. Almheiri, D. Marolf, J. Polchinski, and J. Sully, "Black holes: complementarity or firewalls?," *JHEP* **02** (2013) 062, arXiv:1207.3123.

[19] A. Almheiri, D. Marolf, J. Polchinski, D. Stanford, and J. Sully, "An apologia for firewalls," *JHEP* **09** (2013) 018, arXiv:1304.6483.

[20] C. Cao, S. M. Carroll, and S. Michalakis, "Space from Hilbert space: Recovering geometry from bulk entanglement," *Phys. Rev. D* **95** (2017) 024031, arXiv:1606.08444.

[21] C. Cao and S. M. Carroll, "Bulk entanglement gravity without a boundary: Towards finding Einstein's equation in Hilbert space," arXiv:1712.02803.

[22] K. K. Boddy, S. M. Carroll, and J. Pollack, "How decoherence affects the probability of slow-roll eternal inflation," arXiv:1612.04894.

[23] L. Susskind, L. Thorlacius, and J. Uglum, "The stretched horizon and black hole complementarity," *Phys. Rev. D* **48** (1993) 3743–3761, arXiv:hep-th/9306069.

[24] M. Van Raamsdonk, "Building up spacetime with quantum entanglement," *Gen. Rel. Grav.* **42** (2010) 2323–2329, arXiv:1005.3035. [Int. J. Mod. Phys.D19,2429(2010)].

[25] N. Bao, S. Nezami, H. Ooguri, B. Stoica, J. Sully, and M. Walter, "The holographic entropy cone," *JHEP* **09** (2015) 130, arXiv:1505.07839.

[26] A. Almheiri, X. Dong, and D. Harlow, "Bulk locality and quantum error correction in AdS/CFT," *JHEP* **04** (2015) 163, arXiv:1411.7041.

[27] X. Dong, D. Harlow, and A. C. Wall, "Reconstruction of bulk operators within the entanglement wedge in gauge-gravity duality," *Phys. Rev. Lett.* **117** (2016) 021601, arXiv:1601.05416.

[28] N. Bao and I. F. Halpern, "Holographic inequalities and entanglement of purification," *JHEP* **03** (2018) 006, arXiv:1710.07643.

[29] R. Bousso and N. Engelhardt, "Generalized second law for cosmology," *Phys. Rev. D* **93** (2016) 024025, arXiv:1510.02099.

[30] **Virgo, LIGO Scientific Collaboration**, B. P. Abbott et al., "Observation of gravitational waves from a binary black hole merger," *Phys. Rev. Lett.* **116** (2016) 061102, arXiv:1602.03837.

[31] F. Pastawski, B. Yoshida, D. Harlow, and J. Preskill, "Holographic quantum error-correcting codes: Toy models for the bulk/boundary correspondence," *JHEP* **06** (2015) 149, arXiv:1503.06237.

[32] V. E. Hubeny, M. Rangamani, and T. Takayanagi, "A covariant holographic entanglement entropy proposal," *JHEP* **07** (2007) 062, arXiv:0705.0016.

[33] R. Bousso, Z. Fisher, J. Koeller, S. Leichenauer, and A. C. Wall, "Proof of the quantum null energy condition," *Phys. Rev. D* **93** (2016) 024017, arXiv:1509.02542.

[34] R. Bousso, Z. Fisher, S. Leichenauer, and A. C. Wall, "Quantum focusing conjecture," *Phys. Rev. D* **93** (2016) 064044, arXiv:1506.02669.

[35] A. Singh and S. M. Carroll, *in preparation.*

[36] R. Orus, "A practical introduction to tensor networks: Matrix product states and projected entangled pair states," *Annals Phys.* **349** (2014) 117–158, arXiv:1306.2164.

[37] C. Cheung and G. N. Remmen, "Naturalness and the weak gravity conjecture," *Phys. Rev. Lett.* **113** (2014) 051601, arXiv:1402.2287.

[38] C. Cheung and G. N. Remmen, "Infrared consistency and the weak gravity conjecture," *JHEP* **12** (2014) 087, arXiv:1407.7865.

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

[40] D. N. Page, "Information in black hole radiation," *Phys. Rev. Lett.* **71** (1993) 3743–3746, arXiv:hep-th/9306083.

[41] D. N. Page, "Time dependence of Hawking radiation entropy," *JCAP* **1309** (2013) 028, arXiv:1301.4995.

[42] P. Hayden and J. Preskill, "Black holes as mirrors: Quantum information in random subsystems," *JHEP* **09** (2007) 120, arXiv:0708.4025.

[43] S. W. Hawking, "Particle creation by black holes," *Commun. Math. Phys.* **43** (1975) 199–220.

[44] J. Polchinski, "The black hole information problem," in *Proceedings, Theoretical Advanced Study Institute in Elementary Particle Physics: New Frontiers in Fields and Strings (TASI 2015): Boulder, CO, USA, June 1-26, 2015*, pp. 353–397. 2017. arXiv:1609.04036.

[45] D. Harlow, "Jerusalem lectures on black holes and quantum information," *Rev. Mod. Phys.* **88** (2016) 015002, arXiv:1409.1231.

[46] E. Lubkin, "Entropy of an n-system from its correlation with a k-reservoir," *J. Math. Phys* **19** (1978) 1028.

[47] S. Lloyd and H. Pagels, "Complexity as thermodynamic depth," *Annals Phys.* **188** (1988) 186.

[48] J. Preskill, "Quantum Shannon theory," arXiv:1604.07450.

[49] M. Rangamani and M. Rota, "Entanglement structures in qubit systems," *J. Phys.* **A48** (2015) 385301, arXiv:1505.03696.

[50] P. Hayden, M. Headrick, and A. Maloney, "Holographic mutual information is monogamous," *Phys. Rev. D* **87** (2013) 046003, arXiv:1107.2940.

[51] N. Bao, S. Blitz, and B. Stoica, "Holographic entropy cone measures," arXiv:1701.03498.

[52] J. Cotler, P. Hayden, G. Salton, B. Swingle, and M. Walter, "Entanglement wedge reconstruction via universal recovery channels," arXiv:1704.05839.

[53] S. H. Shenker and D. Stanford, "Black holes and the butterfly effect," *JHEP* **03** (2014) 067, arXiv:1306.0622.

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

[55] T. Banks, M. R. Douglas, G. T. Horowitz, and E. J. Martinec, "AdS dynamics from conformal field theory," arXiv:hep-th/9808016.

[56] A. W. Peet and S. F. Ross, "Microcanonical phases of string theory on AdS(m) x S**n," *JHEP* **12** (1998) 020, arXiv:hep-th/9810200.

[57] S. G. Avery and D. A. Lowe, "Typical event horizons in AdS/CFT," *JHEP* **01** (2016) 082, arXiv:1501.05573.

[58] I. Agol, J. Hass, and W. P. Thurston, "The computational complexity of knot genus and spanning area," *Trans. Amer. Math. Soc.* **358** (2006) 3821, arXiv:math/0205057.

[59] M. Freedman and M. Headrick, "Bit threads and holographic entanglement," *Commun. Math. Phys.* **352** (2017) 407–438, arXiv:1604.00354.

[60] G. Chartrand and O. R. Oellermann, *Applied and algorithmic graph theory.* McGraw-Hill, New York, New York, 1993.

[61] W. Ballmann, *Lectures on spaces of nonpositive curvature.* Birkhäuser, Basel, 1995.

[62] P. Calabrese and J. L. Cardy, "Entanglement entropy and quantum field theory," *J. Stat. Mech.* **0406** (2004) P06002, arXiv:hep-th/0405152.

[63] C. Holzhey, F. Larsen, and F. Wilczek, "Geometric and renormalized entropy in conformal field theory," *Nucl. Phys.* **B424** (1994) 443–467, arXiv:hep-th/9403108.

[64] H. Maxfield, "Entanglement entropy in three dimensional gravity," *JHEP* **04** (2015) 031, arXiv:1412.0687.

[65] N. Lashkari, M. B. McDermott, and M. Van Raamsdonk, "Gravitational dynamics from entanglement 'thermodynamics'," *JHEP* **04** (2014) 195, arXiv:1308.3716.

[66] T. Faulkner, M. Guica, T. Hartman, R. C. Myers, and M. Van Raamsdonk, "Gravitation from entanglement in holographic CFTs," *JHEP* **03** (2014) 051, arXiv:1312.7856.

[67] B. Swingle and M. Van Raamsdonk, "Universality of gravity from entanglement," arXiv:1405.2933.

[68] M. Headrick and V. E. Hubeny, "Riemannian and Lorentzian flow-cut theorems," arXiv:1710.09516.

[69] A. C. Wall, "Maximin surfaces, and the strong subadditivity of the covariant holographic entanglement entropy," *Class. Quant. Grav.* **31** (2014) 225007, arXiv:1211.3494.

[70] M. Banados, C. Teitelboim, and J. Zanelli, "The black hole in three-dimensional space-time," *Phys. Rev. Lett.* **69** (1992) 1849–1851, arXiv:hep-th/9204099.

[71] L. Susskind, "The world as a hologram," *J. Math. Phys.* **36** (1995) 6377–6396, arXiv:hep-th/9409089.

[72] G. 't Hooft, "Dimensional reduction in quantum gravity," arXiv:gr-qc/9310026.

[73] O. Aharony, S. S. Gubser, J. M. Maldacena, H. Ooguri, and Y. Oz, "Large N field theories, string theory and gravity," *Phys. Rept.* **323** (2000) 183–386, arXiv:hep-th/9905111.

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

[75] G. Vidal, "Class of quantum many-body states that can Be efficiently simulated," *Phys. Rev. Lett.* **101** (2008) 110501, quant-ph/0610099.

[76] B. Swingle, "Entanglement renormalization and holography," *Phys. Rev. D* **86** (2012) 065007, arXiv:0905.1317.

[77] X.-L. Qi, "Exact holographic mapping and emergent space-time geometry," arXiv:1309.6282.

[78] B. Czech, L. Lamprou, S. McCandlish, and J. Sully, "Integral geometry and holography," *JHEP* **10** (2015) 175, arXiv:1505.05515.

[79] C. Bény, "Causal structure of the entanglement renormalization ansatz," *New J. Phys.* **15** (Feb., 2013) 023020, arXiv:1110.4872.

[80] G. Evenbly and G. Vidal, "Algorithms for entanglement renormalization," *Phys. Rev. B* **79** (2009) 144108, arXiv:0707.1454.

[81] R. N. C. Pfeifer, G. Evenbly, and G. Vidal, "Entanglement renormalization, scale invariance, and quantum criticality," *Phys. Rev. A* **79** (2009) 040301, arXiv:0810.0580.

[82] G. Evenbly, P. Corboz, and G. Vidal, "Nonlocal scaling operators with entanglement renormalization," *Phys. Rev. B* **82** (Oct., 2010) 132411, arXiv:0912.2166.

[83] Y.-Y. Shi, L.-M. Duan, and G. Vidal, "Classical simulation of quantum many-body systems with a tree tensor network," *Phys. Rev. A* **74** (2006) 022320, quant-ph/0511070.

[84] P. Pfeuty, "The one-dimensional ising model with a transverse field," *Annals Phys.* **57** (1970) 79 – 90.

[85] T. Hartman and J. Maldacena, "Time evolution of entanglement entropy from black hole interiors," *JHEP* **1305** (2013) 014, arXiv:1303.1080.

[86] B. Swingle, "Constructing holographic spacetimes using entanglement renormalization," arXiv:1209.3304.

[87] G. Evenbly and G. Vidal, "Scaling of entanglement entropy in the (branching) multiscale entanglement renormalization ansatz," *Phys. Rev. B* **89** (2014) 235113, arXiv:1310.8372.

[88] J. Haegeman, T. J. Osborne, H. Verschelde, and F. Verstraete, "Entanglement renormalization for quantum fields in real space," *Phys. Rev. Lett.* **110** (2013) 100402, arXiv:1102.5524.

[89] M. Nozaki, S. Ryu, and T. Takayanagi, "Holographic geometry of entanglement renormalization in quantum field theories," *JHEP* **10** (2012) 193, arXiv:1208.3469.

[90] A. Mollabashi, M. Nozaki, S. Ryu, and T. Takayanagi, "Holographic geometry of cMERA for quantum quenches and finite temperature," *JHEP* **1403** (2014) 098, arXiv:1311.6095.

[91] J. Brown and M. Henneaux, "Central charges in the canonical realization of asymptotic symmetries: An example from three dimensional gravity," *Commun. Math. Phys.* **104** (1986) 207–226.

[92] S. Ryu and T. Takayanagi, "Aspects of holographic entanglement entropy," *JHEP* **08** (2006) 045, arXiv:hep-th/0605073.

[93] T. Jacobson, "Thermodynamics of space-time: The Einstein equation of state," *Phys. Rev. Lett.* **75** (1995) 1260–1263, arXiv:gr-qc/9504004.

[94] J. D. Bekenstein, "Black holes and the second law," *Lettere al Nuovo Cimento (1971-1985)* **4** (1972) 737–740.

[95] R. Bousso, H. Casini, Z. Fisher, and J. Maldacena, "Proof of a quantum Bousso bound," *Phys. Rev. D* **90** (2014) 044002, arXiv:1404.5635.

[96] R. Bousso, H. Casini, Z. Fisher, and J. Maldacena, "Entropy on a null surface for interacting quantum field theories and the Bousso bound," arXiv:1406.4545.

[97] R. Bousso, "The holographic principle," *Rev. Mod. Phys.* **74** (2002) 825–874, arXiv:hep-th/0203101.

[98] C. Martinez, C. Teitelboim, and J. Zanelli, "Charged rotating black hole in three space-time dimensions," *Phys. Rev. D* **61** (2000) 104013, arXiv:hep-th/9912259.

[99] M. Srednicki, "Entropy and area," *Phys. Rev. Lett.* **71** (1993) 666–669, arXiv:hep-th/9303048.

[100] G. Evenbly and G. Vidal, "Tensor network states and geometry," *J. Stat. Phys.* **145** (Nov., 2011) 891–918, arXiv:1106.1082.

[101] G. Evenbly and G. Vidal, "Tensor network renormalization yields the multi-scale entanglement renormalization ansatz," 1502.05385.

[102] J. Molina-Vilaplana and J. Prior, "Entanglement, tensor networks and black hole horizons," *Gen. Rel. Grav.* **46** (2014) 1823, arXiv:1403.5395.

[103] M. Heydeman, M. Marcolli, I. Saberi, and B. Stoica, "Tensor networks, p-adic fields, and algebraic curves: arithmetic and the $\mathrm{AdS}_3/\mathrm{CFT}_2$ correspondence," arXiv:1605.07639.

[104] S. S. Gubser, M. Heydeman, C. Jepsen, M. Marcolli, S. Parikh, I. Saberi, B. Stoica, and B. Trundy, "Edge length dynamics on graphs with applications to p-adic AdS/CFT," arXiv:1612.09580.

[105] B. Czech, L. Lamprou, S. McCandlish, and J. Sully, "Tensor networks from kinematic space," *JHEP* **07** (2016) 100, arXiv:1512.01548.

[106] B. Czech, G. Evenbly, L. Lamprou, S. McCandlish, X.-L. Qi, J. Sully, and G. Vidal, "Tensor network quotient takes the vacuum to the thermal state," *Phys. Rev. B* **94** (2016) 085101, arXiv:1510.07637.

[107] G. Evenbly, "Hyperinvariant tensor networks and holography," *Phys. Rev. Lett.* **119** (2017) 141602, arXiv:1704.04229.

[108] A. Strominger, "The dS / CFT correspondence," *JHEP* **10** (2001) 034, arXiv:hep-th/0106113.

[109] Y. Sekino and L. Susskind, "Census taking in the hat: FRW/CFT duality," *Phys. Rev. D* **80** (2009) 083531, arXiv:0908.3844.

[110] R. Sinai Kunkolienkar and K. Banerjee, "Towards a dS/MERA correspondence," arXiv:1611.08581.

[111] B. Czech, L. Lamprou, S. McCandlish, B. Mosk, and J. Sully, "A stereoscopic look into the bulk," *JHEP* **07** (2016) 129, arXiv:1604.03110.

[112] N. Bao, C. Cao, S. M. Carroll, and L. McAllister, "Quantum circuit cosmology: The expansion of the universe since the first qubit," arXiv:1702.06959.

[113] Y. Nomura, "Physical theories, eternal inflation, and quantum universe," *JHEP* **11** (2011) 063, arXiv:1104.2324.

[114] Y. Nomura, "Quantum mechanics, spacetime locality, and gravity," *Found. Phys.* **43** (2013) 978–1007, arXiv:1110.4630.

[115] A. R. Brown, D. A. Roberts, L. Susskind, B. Swingle, and Y. Zhao, "Holographic complexity equals bulk action?," *Phys. Rev. Lett.* **116** (2016) 191301, arXiv:1509.07876.

[116] G. W. Gibbons and S. W. Hawking, "Cosmological event horizons, thermodynamics, and particle creation," *Phys. Rev. D* **15** (1977) 2738–2751.

[117] M. Miyaji, T. Takayanagi, and K. Watanabe, "From path integrals to tensor networks for the AdS/CFT correspondence," *Phys. Rev. D* **95** (2017) 066004, arXiv:1609.04645.

[118] R. N. C. Pfeifer, G. Evenbly, and G. Vidal, "Entanglement renormalization, scale invariance, and quantum criticality," *Phys. Rev. A* **79** (Apr., 2009) 040301, arXiv:0810.0580.

[119] R. M. Wald, "Asymptotic behavior of homogeneous cosmological models in the presence of a positive cosmological constant," *Phys. Rev. D* **28** (1983) 2118–2120.

[120] A. A. Starobinskii, "Isotropization of arbitrary cosmological expansion given an effective cosmological constant," *JETP Lett.* **37** (1983) 66.

[121] J. D. Barrow, "Cosmic no hair theorems and inflation," *Phys. Lett. B* **187** (1987) 12–16.

[122] J. D. Barrow and G. Goetz, "The asymptotic approach to de Sitter space-time," *Phys. Lett. B* **231** (1989) 228–230.

[123] Y. Kitada and K.-i. Maeda, "Cosmic no hair theorem in power law inflation," *Phys. Rev. D* **45** (1992) 1416–1419.

[124] Y. Kitada and K.-i. Maeda, "Cosmic no hair theorem in homogeneous space-times. 1. Bianchi models," *Class. Quant. Grav.* **10** (1993) 703–734.

[125] M. Bruni, S. Matarrese, and O. Pantano, "A local view of the observable universe," *Phys. Rev. Lett.* **74** (1995) 1916–1919, arXiv:astro-ph/9407054.

[126] M. Bruni, F. C. Mena, and R. K. Tavakol, "Cosmic no hair: Nonlinear asymptotic stability of de Sitter universe," *Class. Quant. Grav.* **19** (2002) L23–L29, arXiv:gr-qc/0107069.

[127] W. Boucher and G. W. Gibbons, "Cosmic baldness," in *Nuffield Workshop on the Very Early Universe Cambridge, England, June 21-July 9, 1982*, pp. 273–278. 2011. arXiv:1109.3535.

[128] A. Maleknejad and M. M. Sheikh-Jabbari, "Revisiting cosmic no-hair theorem for inflationary settings," *Phys. Rev. D* **85** (2012) 123508, arXiv:1203.0219.

[129] S. Hollands, "Correlators, Feynman diagrams, and quantum no-hair in de Sitter spacetime," *Commun. Math. Phys.* **319** (2013) 1–68, arXiv:1010.5367.

[130] D. Marolf and I. A. Morrison, "The IR stability of de Sitter: Loop corrections to scalar propagators," *Phys. Rev. D* **82** (2010) 105032, arXiv:1006.0035.

[131] D. Marolf and I. A. Morrison, "The IR stability of de Sitter QFT: results at all orders," *Phys. Rev. D* **84** (2011) 044040, arXiv:1010.5327.

[132] M. Raginsky, "Strictly contractive quantum channels and physically realizable quantum computers," *Phys. Rev. A* **65** (Mar., 2002) 032306, quant-ph/0105141.

[133] K. S. Thorne, R. H. Price, and D. A. Macdonald, *Black holes: The membrane paradigm*. Yale University Press, New Haven, Connecticut, 1986.

[134] W. Fischler, "Taking de Sitter seriously." Talk given at Role of Scaling Laws in Physics and Biology (Celebrating the 60th Birthday of Geoffrey West), Santa Fe, Dec., 2000.

[135] T. Banks, "QuantuMechanics and CosMology." Talk given at the festschrift for L. Susskind, Stanford University, May 2000, 2000.

[136] R. Bousso, "Positive vacuum energy and the n-bound," *JHEP* **2000** (2000) 038, https://arxiv.org/abs/hep-th/0010252.

[137] T. Banks, "Cosmological breaking of supersymmetry?," *Int. J. Mod. Phys. A* **16** (2001) 910–921, arXiv:hep-th/0007146.

[138] E. Witten, "Quantum gravity in de Sitter space," in *Strings 2001: International Conference Mumbai, India, January 5-10, 2001*. 2001. arXiv:hep-th/0106109.

[139] L. Dyson, M. Kleban, and L. Susskind, "Disturbing implications of a cosmological constant," *JHEP* **10** (2002) 011, arXiv:hep-th/0208013.

[140] M. K. Parikh and E. P. Verlinde, "De Sitter holography with a finite number of states," *JHEP* **01** (2005) 054, arXiv:hep-th/0410227.

[141] A. Albrecht and L. Sorbo, "Can the universe afford inflation?," *Phys. Rev. D* **70** (2004) 063528, arXiv:hep-th/0405270.

[142] S. M. Carroll, "Why Boltzmann brains are bad," arXiv:1702.00850.

[143] K. K. Boddy, S. M. Carroll, and J. Pollack, "De Sitter space without dynamical quantum fluctuations," *Found. Phys.* **46** (2016) 702–735, arXiv:1405.0298.

[144] T. Banks and W. Fischler, "Holographic cosmology 3.0," *Phys. Scripta* **T117** (2005) 56–63, arXiv:hep-th/0310288.

[145] T. Banks, "Holographic space-time: The takeaway," arXiv:1109.2435.

[146] T. Banks and W. Fischler, "Holographic inflation revised," arXiv:1501.01686.

[147] D. Petz, "Sufficient subalgebras and the relative entropy of states of a von neumann algebra," *Commun. Math. Phys.* **105** (1986) 123–131.

[148] D. Petz, "Sufficiency of channels over von neumann algebras," *Q. J. Math.* **39** (1988) 97–108.

[149] D. Petz, "Monotonicity of quantum relative entropy revisited," *Rev. Math. Phys.* **15** (2003) 79–91.

[150] M. Junge, R. Renner, D. Sutter, M. M. Wilde, and A. Winter, "Universal recovery from a decrease of quantum relative entropy," arXiv:1509.07127.

[151] Á. Rivas and S. F. Huelga, *Open quantum systems: An introduction*. Springer, Berlin, 2012.

[152] M. A. Nielsen and I. L. Chuang, *Quantum computation and quantum information: 10th Anniversary Edition*. Cambridge University Press, Cambridge, 10th ed., 2011.

[153] J. Preskill, "Lecture notes for physics 219: Quantum computation," 1997. http://www.theory.caltech.edu/people/preskill/ph229/notes/chap7.pdf.

[154] A. Reynolds and S. F. Ross, "Complexity in de Sitter space," *Class. Quant. Grav.* **34** (2017) 175013, arXiv:1706.03788.

[155] R. A. Jefferson and R. C. Myers, "Circuit complexity in quantum field theory," arXiv:1707.08570.

[156] S. Chapman, M. P. Heller, H. Marrochio, and F. Pastawski, "Towards complexity for quantum field theory states," arXiv:1707.08582.

[157] D. Harlow, S. H. Shenker, D. Stanford, and L. Susskind, "Tree-like structure of eternal inflation: A solvable model," *Phys. Rev. D* **85** (2012) 063516, arXiv:1110.0496.

[158] G. Evenbly and G. Vidal, "Entanglement renormalization in two spatial dimensions," *Phys. Rev. Lett.* **102** (May, 2009) 180406, arXiv:0811.0879.

[159] D. Kastor, S. Ray, and J. Traschen, "Genuine cosmic hair," *Class. Quant. Grav.* **34** (2017) 045003, arXiv:1608.04641.

[160] R. M. Wald, *General relativity*. University of Chicago Press, Chicago, Illinois, 1984.

[161] B. K. Berger, *Singularities in cosmological spacetimes*, pp. 437–460. Springer, Berlin, 2014.

[162] E. Kolb and M. Turner, *The early universe*. Westview Press, Boulder, Colorado, 1994.

[163] H. J. Schmidt, "De Sitter space-time as an attractor solution in fourth order gravity," *Class. Quant. Grav.* **5** (1988) 233.

[164] U. Brauer, A. Rendall, and O. Reula, "The cosmic no hair theorem and the nonlinear stability of homogeneous Newtonian cosmological models," *Class. Quant. Grav.* **11** (1994) 2283–2296, arXiv:gr-qc/9403050.

[165] M. Kleban and L. Senatore, "Inhomogeneous anisotropic cosmology," *JCAP* **1610** (2016) 022, arXiv:1602.03520.

[166] M. Spradlin, A. Strominger, and A. Volovich, "Les Houches lectures on de Sitter space," in *Unity from duality: Gravity, gauge theory and strings. Proceedings, NATO Advanced Study Institute, Euro Summer School, 76th session, Les Houches, France, July 30-August 31, 2001*, pp. 423–453. 2001. arXiv:hep-th/0110007.

[167] S. M. Carroll and J. Chen, "Spontaneous inflation and the origin of the arrow of time," arXiv:hep-th/0410270.

[168] A. Albrecht, "De Sitter equilibrium as a fundamental framework for cosmology," *J. Phys. Conf. Ser.* **174** (2009) 012006, arXiv:0906.1047.

[169] P. B. Krishna and T. K. Mathew, "Holographic equipartition and the maximization of entropy," *Phys. Rev. D* **96** (2017) 063513, arXiv:1702.02787.

[170] T. Padmanabhan, "Thermodynamical aspects of gravity: New insights," *Rept. Prog. Phys.* **73** (2010) 046901, arXiv:0911.5004.

[171] E. P. Verlinde, "On the origin of gravity and the laws of Newton," *JHEP* **04** (2011) 029, arXiv:1001.0785.

[172] T. Jacobson, "Entanglement equilibrium and the Einstein equation," *Phys. Rev. Lett.* **116** (2016) 201101, arXiv:1505.04753.

[173] S. M. Carroll and G. N. Remmen, "What is the entropy in entropic gravity?," *Phys. Rev. D* **93** (2016) 124052, arXiv:1601.07558.

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

[175] R. Bousso, "Holography in general space-times," *JHEP* **06** (1999) 028, arXiv:hep-th/9906022.

[176] S. A. Hayward, "General laws of black-hole dynamics," *Phys. Rev. D* **49** (Jun, 1994) 6467–6474, arXiv:gr-qc/9303006.

[177] A. Ashtekar and B. Krishnan, "Dynamical horizons: Energy, angular momentum, fluxes and balance laws," *Phys. Rev. Lett.* **89** (2002) 261101, arXiv:gr-qc/0207080.

[178] A. Ashtekar and B. Krishnan, "Dynamical horizons and their properties," *Phys. Rev. D* **68** (2003) 104030, arXiv:gr-qc/0308033.

[179] A. Ashtekar and B. Krishnan, "Isolated and dynamical horizons and their applications," *Living Rev. Rel.* **7** (2004) 10, arXiv:gr-qc/0407042.

[180] I. Booth, "Black hole boundaries," *Can. J. Phys.* **83** (2005) 1073–1099, arXiv:gr-qc/0508107.

[181] I. Booth, L. Brits, J. A. Gonzalez, and C. Van Den Broeck, "Marginally trapped tubes and dynamical horizons," *Class. Quant. Grav.* **23** (2006) 413–440, arXiv:gr-qc/0506119.

[182] R. Bousso and N. Engelhardt, "New area law in general relativity," *Phys. Rev. Lett.* **115** (2015) 081301, arXiv:1504.07627.

[183] R. Bousso and N. Engelhardt, "Proof of a new area law in general relativity," *Phys. Rev. D* **92** (2015) 044031, arXiv:1504.07660.

[184] F. Sanches and S. J. Weinberg, "Holographic entanglement entropy conjecture for general spacetimes," *Phys. Rev. D* **94** (2016) 084034, arXiv:1603.05250.

[185] J. M. Bardeen, B. Carter, and S. W. Hawking, "The four laws of black hole mechanics," *Commun. Math. Phys.* **31** (1973) 161–170.

[186] T. Jacobson, "Introductory lectures on black hole thermodynamics," *Given at Utrecht U. in* (1996) 30–59.

[187] S. W. Hawking, "Black holes in general relativity," *Commun. Math. Phys.* **25** (1972) 152–166.

[188] P. T. Saunders, "Observations in some simple cosmological models with shear," *Monthly Notices of the Royal Astronomical Society* **142** (1969) 213–227.

[189] P. Fleury, F. Nugier, and G. Fanizza, "Geodesic-light-cone coordinates and the Bianchi I spacetime," *JCAP* **1606** (2016) 008, arXiv:1602.04461.

[190] A. Chatwin-Davies, "Companion calculations for 'cosmic equilibration: A holographic no-hair theorem from the generalized second law'." https://doi.org/10.5281/zenodo.1184720, 2018.

[191] V. F. Mukhanov, H. A. Feldman, and R. H. Brandenberger, "Theory of cosmological perturbations. Part 1. Classical perturbations. Part 2. Quantum theory of perturbations. Part 3. Extensions," *Phys. Rept.* **215** (1992) 203–333.

[192] N. D. Birrell and P. C. W. Davies, *Quantum fields in curved space*. Cambridge University Press, Cambridge, 1984.

[193] S. Hossenfelder, "Minimal length scale scenarios for quantum gravity," *Living Rev. Rel.* **16** (2013) 2, arXiv:1203.6191.

[194] C. Rovelli, "Strings, loops and others: a critical survey of the present approaches to quantum gravity," arXiv:gr-qc/9803024v3.

[195] S. Carlip, D.-W. Chiou, W.-T. Ni, and R. Woodard, "Quantum gravity: A brief history of ideas and some prospects," *Int. J. Mod. Phys. D* **24** (2015) 1530028, arXiv:1507.08194.

[196] T. Jacobson, "Trans-planckian redshifts and the substance of the space-time river," *Prog. Theor. Phys. Supp.* **136** (1999) 1–17, arXiv:1507.08194.

[197] R. H. Brandenberger, "Principles, progress, and problems in inflationary cosmology," *AAPPS Bulletin* **11** (2001) 20–29, arXiv:astro-ph/0208103.

[198] A. Kempf, "Mode generating mechanism in inflation with a cutoff," *Phys. Rev. D* **63** (2001) 083514, arXiv:astro-ph/0009209.

[199] J. Martin and R. H. Brandenberger, "Trans-planckian problem of inflationary cosmology," *Phys. Rev. D* **63** (2001) 123501, arXiv:hep-th/0005209.

[200] A. Kempf and J. C. Niemeyer, "Perturbation spectrum in inflation with a cutoff," *Phys. Rev. D* **64** (2001) 103501, arXiv:astro-ph/0103225.

[201] G. Shiu, "Inflation as a probe of trans-Planckian physics: a brief review and progress report," *JPCS* **18** (2005) 188.

[202] T. Padmanabhan, "Acceptable density perturbations from inflation due to quantum gravitational damping," *Phys. Rev. Lett.* **60** (1988) 2229–2230.

[203] T. Padmanabhan, T. R. Seshadri, and T. P. Singh, "Making inflation work: Damping of density perturbations due to planck energy cutoff," *Phys. Rev. D* **39** (1989) 2100–2107.

[204] R. H. Brandenberger and J. Martin, "The robustness of inflation to changes in super-planck-scale physics," *Mod. Phys. Lett. A* **16** (2001) 999–1006, arXiv:astro-ph/0005432.

[205] R. H. Brandenberger and J. Martin, "On signatures of short distance physics in the cosmic microwave background," *Int. J. Mod. Phys.* **A17** (2002) 3663–3680, arXiv:hep-th/0202142.

[206] R. Easther, B. R. Greene, W. H. Kinney, and G. Shiu, "Inflation as a probe of short distance physics," *Phys. Rev. D* **64** (2001) 103502, arXiv:hep-th/0104102.

[207] J. C. Niemeyer, "Inflation with a planck-scale frequency cutoff," *Phys. Rev. D* **63** (2001) 123502, arXiv:astro-ph/0005533.

[208] R. Easther, B. R. Greene, W. H. Kinney, and G. Shiu, "Generic estimate of trans-planckian modifications to the primordial power spectrum in inflation," *Phys. Rev. D* **66** (2002) 023518, arXiv:hep-th/0204129.

[209] R. Easther, B. R. Greene, W. H. Kinney, and G. Shiu, "Imprints of short distance physics on inflationary cosmology," *Phys. Rev. D* **67** (2003) 063508, arXiv:hep-th/0110226.

[210] R. H. Brandenberger and J. Martin, "Back-reaction and the trans-planckian problem of inflation reexamined," *Phys. Rev. D* **71** (2005) 023504, arXiv:hep-th/0410223.

[211] R. Easther, W. H. Kinney, and H. Peiris, "Boundary effective field theory and trans-planckian perturbations: astrophysical implications," *J. Cosmol. Astropart. Phys.* **2005** (2005) 001, arXiv:astro-ph/0505426.

[212] B. R. Greene, K. Schalm, G. Shiu, and J. P. van der Schaar, "Decoupling in an expanding universe: backreaction barely constrains short distance effects in the cosmic microwave background," *J. Cosmol. Astropart. Phys.* **2005** (2005) 001, arXiv:hep-th/0411217.

[213] L. Sriramkumar and T. Padmanabhan, "Initial state of matter fields and trans-planckian physics: Can cmb observations disentangle the two?," *Phys. Rev. D* **71** (2005) 103512, arXiv:gr-qc/0408034.

[214] U. H. Danielsson, "Note on inflation and trans-planckian physics," *Phys. Rev. D* **66** (2002) 023511, arXiv:hep-th/0203198.

[215] D. Polarski and A. A. Starobinsky, "Semiclassicality and decoherence of cosmological perturbations," *Class. Quant. Grav.* **13** (1996) 377–392, arXiv:gr-qc/9504030.

[216] C. Kiefer, D. Polarski, and A. A. Starobinsky, "Quantum-to-classical transition for fluctuations in the early universe," *Int. J. Mod. Phys. D* **07** (1998) 455–462, arXiv:gr-qc/9802003.

[217] C. Kiefer and D. Polarski, "Why do cosmological perturbations look classical to us?," *Adv. Sci. Lett.* **2** (2009) 164–173, arXiv:0810.0087.

[218] E. Nelson, "Quantum decoherence during inflation from gravitational nonlinearities," *J. Cosmol. Astropart. Phys.* **2016** (2016) 22, arXiv:1601.03734.

[219] E. Joos, H. D. Zeh, C. Kiefer, D. J. W. Giulini, J. Kupsch, and I.-O. Stamatescu, *Decoherence and the appearance of a classical world in quantum theory*. Springer-Verlag, Berlin, 2003.

[220] C. Shannon, *The mathematical theory of communication*. University of Illinois Press, Champaign, Illinois, 1949.

[221] H. J. Landau, "Necessary density conditions for sampling and interpolation of certain entire functions," *Acta Mathematica* **117** (1967) 37–52.

[222] A. Jerri, "The Shannon sampling theorem - its various extensions and applications: A tutorial review," *Proc. IEEE* **65** (1977) 1565–1596.

[223] A. Kempf, A. Chatwin-Davies, and R. T. W. Martin, "A fully covariant information-theoretic ultraviolet cutoff for scalar fields in expanding Friedmann Robertson Walker spacetimes," *J. Math. Phys.* **54** (2013) 022301, arXiv:1210.0750.

[224] A. Kempf and R. Martin, "Information theory, spectral geometry, and quantum gravity," *Phys. Rev. Lett.* **100** (2008) 021304, arXiv:0708.0062.

[225] M. A. Naimark, *Linear differential operators, part II*. Frederick Ungar Publishing Company, New York, New York, 1968.

[226] W. Amrein, A. M. Hinz, and D. Pearson, eds., *Sturm-Liouville theory: past and present*. Springer Science & Business Media, Berlin, 2005.

[227] A. Zettl, *Sturm-Liouville theory*. Mathematical Surveys and Monographs. American Mathematical Society, Providence, Rhode Island, 2005.

[228] D. N. Page, "Is black hole evaporation predictable?," *Phys. Rev. Lett.* **44** (1980) 301.

[229] V. Mukhanov and S. Winitzki, *Introduction to quantum effects in gravity*. Cambridge University Press, Cambridge, 2007.

[230] D. N. Page, "Particle emission rates from a black hole: Massless particles from an uncharged, nonrotating hole," *Phys. Rev. D* **13** (1976) 198–206.

[231] D. N. Page, "Particle emission rates from a black hole. 2. Massless particles from a rotating hole," *Phys. Rev. D* **14** (1976) 3260–3273.

[232] Chatwin-Davies, Aidan, "A covariant natural ultraviolet cutoff in inflationary cosmology," Master's thesis, University of Waterloo, 2013. http://hdl.handle.net/10012/7759.

[233] S. W. Hawking, "Information loss in black holes," *Phys. Rev. D* **72** (2005) 084013, arXiv:hep-th/0507171.

[234] L. Susskind and J. Lindesay, *An introduction to black holes, information and the string theory revolution: The holographic universe*. World Scientific, Singapore, 2005.

[235] S. D. Mathur, "The information paradox: a pedagogical introduction," *Class. Quantum Grav.* **26** (2009) 224001, arXiv:0909.1038.

[236] V. Balasubramanian, D. Marolf, and M. Rozali, "Information recovery from black holes," *Gen. Rel. Grav.* **38** (2006) 1529–1536, arXiv:hep-th/0604045.

[237] D. Marolf, "Unitarity and holography in gravitational physics," *Phys. Rev. D* **79** (2009) 044010, arXiv:0808.2842.

[238] P. Hayden and J. Preskill, "Black holes as mirrors: quantum information in random subsystems," *JHEP* **2007** (2007) 120, arXiv:0708.4025.

[239] A. Messiah, *Quantum mechanics volume II*. North-Holland, Amsterdam, 1965.

[240] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters, "Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels," *Phys. Rev. Lett.* **70** (1993) 1895–1899.

[241] M. Wilde, *Quantum information theory*. Cambridge University Press, Cambridge, 2013.

[242] S. H. Shenker and D. Stanford, "Stringy effects in scrambling," *JHEP* **05** (2015) 132, arXiv:1412.6087.

[243] J. Maldacena, S. H. Shenker, and D. Stanford, "A bound on chaos," *JHEP* **08** (2016) 106, arXiv:1503.01409.

[244] P. Chen, Y. C. Ong, and D.-h. Yeom, "Black hole remnants and the information loss paradox," *Phys. Rept.* **603** (2015) 1, arXiv:1412.8366.

[245] S. Lloyd and J. Preskill, "Unitarity of black hole evaporation in final-state projection models," *JHEP* **08** (2014) 126, arXiv:1308.4209.

[246] K. Papadodimas and S. Raju, "An infalling observer in AdS/CFT," *JHEP* **10** (2013) 212, arXiv:1211.6767.

[247] S. D. Mathur, "The fuzzball proposal for black holes: an elementary review," *Fortsch. Phys.* **53** (2005) 793, arXiv:hep-th/0502050.

[248] S. B. Giddings, "Nonviolent nonlocality," *Phys. Rev. D* **88** (2013) 064023, arXiv:1211.7070.

[249] M. Hotta and A. Sugita, "The fall of black hole firewall: natural nonmaximal entanglement for Page curve," *PTEP* **2015** (2015) 123B04, arXiv:1505.05870.

[250] Y. Nomura, F. Sanches, and S. J. Weinberg, "Black hole interior in quantum gravity," *Phys. Rev. Lett.* **114** (2015) 201301, arXiv:1412.7539.

[251] Y. Nomura, F. Sanches, and S. J. Weinberg, "Relativeness in quantum gravity: limitations and frame dependence of semiclassical descriptions," *JHEP* **04** (2015) 158, arXiv:1412.7538.

[252] Y. Nomura and N. Salzetta, "Why firewalls need not exist," *Phys. Lett. B* **761** (2016) 62–69, arXiv:1602.07673.

[253] D. Harlow and P. Hayden, "Quantum computation vs. firewalls," *JHEP* **06** (2013) 085, arXiv:1301.4504.

[254] C. H. Bennett, H. J. Bernstein, S. Popescu, and B. Schumacher, "Concentrating partial entanglement by local operations," *Phys. Rev. A* **53** (1996) 2046–2052, arXiv:quant-ph/9511030.

[255] J. Oppenheim and W. G. Unruh, "Firewalls and flat mirrors: An alternative to the AMPS experiment which evades the Harlow-Hayden obstacle," *JHEP* **03** (2014) 120, arXiv:1401.1523.

[256] S. Aaronson, "Computation complexity underpinnings of the Harlow-Hayden argument." http://online.kitp.ucsb.edu/online/fuzzorfire_m13/aaronson/.

[257] Y. Sekino and L. Susskind, "Fast scramblers," *JHEP* **10** (2008) 065, arXiv:0808.2096.

[258] N. Lashkari, D. Stanford, M. Hastings, T. Osborne, and P. Hayden, "Towards the fast scrambling conjecture," *JHEP* **04** (2013) 022, arXiv:1111.6580.

[259] S. B. Giddings, "Hawking radiation, the Stefan–Boltzmann law, and unitarization," *Phys. Lett. B* **754** (2016) 39–42, arXiv:1511.08221.

[260] F. G. S. L. Brandão, A. W. Harrow, and M. Horodecki, "Local random quantum circuits are approximate polynomial-fesigns," *Commun. Math. Phys.* **346** (2016) 397–434, arXiv:1208.0692.

[261] C. Dankert, R. Cleve, J. Emerson, and E. Livine, "Exact and approximate unitary 2-designs and their application to fidelity estimation," *Phys. Rev. A* **80** (2009) 012304.

[262] J. Shore and R. Johnson, "Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy," *IEEE Trans. Inf. Theory* **26** (1980) 26–37.

[263] B. Freivogel, R. Jefferson, L. Kabir, B. Mosk, and I.-S. Yang, "Casting shadows on holographic reconstruction," *Phys. Rev. D* **91** (2015) 086013, arXiv:1412.5175.

[264] S. Raju, "Smooth causal patches for AdS black holes," *Phys. Rev. D* **95** (2017) 126002, arXiv:1604.03095.

[265] S. D. Mathur and D. Turton, "The flaw in the firewall argument," *Nucl. Phys. B* **884** (2014) 566–611, arXiv:1306.5488.

[266] S. D. Mathur, "A model with no firewall," arXiv:1506.04342.

[267] G. Dotti and R. J. Gleiser, "Gravitational instability of the inner static region of a Reissner-Nordstrom black hole," *Class. Quant. Grav.* **27** (2010) 185007, arXiv:1001.0152.

[268] N. Engelhardt and G. T. Horowitz, "Holographic consequences of a no transmission principle," *Phys. Rev. D* **93** (2016) 026005, arXiv:1509.07509.

[269] D. Wallace, "Why black hole information loss is paradoxical," arXiv:1710.03783.

[270] E. H. Lieb and M. B. Ruskai, "A fundamental property of quantum-mechanical entropy," *Phys. Rev. Lett.* **30** (1973) 434.

[271] W. G. Unruh and R. M. Wald, "Information loss," *Rept. Prog. Phys.* **80** (2017) 092002, arXiv:1703.02140.

[272] T. Hertog and J. Hartle, "Observational implications of fuzzball formation," arXiv:1704.02123.

[273] N. Bao, A. Bouland, and S. P. Jordan, "Grover search and the no-signaling principle," *Phys. Rev. Lett.* **117** (2016) 120501, arXiv:1511.00657.

[274] K. Osuga and D. N. Page, "Qubit transport model for unitary black hole evaporation without firewalls," arXiv:1607.04642.

[275] S. D. H. Hsu, "Macroscopic superpositions and black hole unitarity," arXiv:1302.0451.

[276] S. D. H. Hsu, "Factorization of unitarity and black hole firewalls," arXiv:1308.5686.

[277] J. Hartle and T. Hertog, "Quantum transitions between classical histories," *Phys. Rev. D* **92** (2015) 063509, arXiv:1502.06770.

[278] K. Papadodimas and S. Raju, "Remarks on the necessity and implications of state-dependence in the black hole interior," *Phys. Rev. D* **93** (2016) 084049, arXiv:1503.08825.

[279] Y. Nomura, J. Varela, and S. J. Weinberg, "Black holes or firewalls: a theory of horizons," *Phys. Rev. D* **88** (2013) 084052, arXiv:1308.4121.

[280] E. Verlinde and H. Verlinde, "Passing through the firewall," arXiv:1306.0515.

[281] Y. Nomura, J. Varela, and S. J. Weinberg, "Complementarity endures: no firewall for an infalling observer," *JHEP* **03** (2013) 059, arXiv:1207.6626.

[282] Y. Nomura and J. Varela, "A note on (no) firewalls: the entropy argument," *JHEP* **07** (2013) 124, arXiv:1211.7033.

[283] D.-h. Yeom, "Information loss problem and roles of instantons," in *Proceedings, 2nd LeCosPA Symposium: Everything about Gravity, Celebrating the Centenary of Einstein's General Relativity (LeCosPA2015): Taipei, Taiwan, December 14-18, 2015*, p. 566. 2017. arXiv:1601.02366.

[284] R. Bousso, "Firewalls from double purity," *Phys. Rev. D* **88** (2013) 084035, arXiv:1308.2665.

[285] S. B. Giddings and R. A. Porto, "The gravitational S-matrix," *Phys. Rev. D* **81** (2010) 025002, arXiv:0908.0004.

[286] A. Strominger, "On BMS invariance of gravitational scattering," *JHEP* **07** (2014) 152, arXiv:1312.2229.

[287] T. Maudlin, "(Information) paradox lost," arXiv:1705.03541.

[288] A. L. Fitzpatrick and J. Kaplan, "Unitarity and the holographic S-matrix," *JHEP* **10** (2012) 032, arXiv:1112.4845.

[289] B. S. Dewitt, "Quantum theory of gravity. I. the canonical theory," *Phys. Rev.* **160** (1967) 1113.

[290] H. D. Zeh, "On the interpretation of measurement in quantum theory," *Found. Phys.* **1** (1970) 69.

[291] W. H. Zurek, "Pointer basis of quantum apparatus: into what mixture does the wave packet collapse?," *Phys. Rev. D* **24** (1981) 1516.

[292] R. B. Griffiths, "Consistent histories and the interpretation of quantum mechanics," *J. Stat. Phys.* **36** (1984) 219.

[293] E. Joos and H. D. Zeh, "The emergence of classical properties through interaction with the environment," *Z. Phys. B* **59** (1985) 223.

[294] M. Schlosshauer, "Decoherence, the measurement problem, and interpretations of quantum mechanics," *Rev. Mod. Phys.* **76** (2004) 1267, arXiv:quant-ph/0312059.

[295] D. Harlow, "The Ryu-Takayanagi formula from quantum error correction," *Commun. Math. Phys.* **354** (2017) 865, arXiv:1607.03901.

[296] D. Carney, L. Chaurette, D. Neuenfeld, and G. W. Semenoff, "Infrared quantum information," *Phys. Rev. Lett.* **119** (2017) 180502, arXiv:1706.03782.

[297] D. Carney, L. Chaurette, D. Neuenfeld, and G. W. Semenoff, "Dressed infrared quantum information," *Phys. Rev. D* **97** (2018) 025007, arXiv:1710.02531.

[298] D. Carney, L. Chaurette, D. Neuenfeld, and G. Semenoff, "On the need for soft dressing," arXiv:1803.02370.

[299] A. Strominger and A. Zhiboedov, "Gravitational memory, BMS supertranslations and soft theorems," *JHEP* **01** (2016) 086, arXiv:1411.5745.

[300] S. W. Hawking, M. J. Perry, and A. Strominger, "Soft hair on black holes," *Phys. Rev. Lett.* **116** (2016) 231301, arXiv:1601.00921.

[301] A. Strominger, "Black hole information revisited," arXiv:1706.07143.

[302] D. Marolf and A. C. Wall, "Eternal black holes and superselection in AdS/CFT," *Class. Quant. Grav.* **30** (2013) 025001, arXiv:1210.3590.

[303] J. Pollack and A. Singh, "Towards space from Hilbert space: Finding lattice structure in finite-dimensional quantum systems," arXiv:1801.10168.

[304] A. S. Holevo, "The capacity of the quantum channel with general signal states," *IEEE Trans. Inf. Theory* **44** (1998) 269, arXiv:quant-ph/9611023.

[305] N. Bao and H. Ooguri, "Distinguishability of black hole microstates," *Phys. Rev. D* **96** (2017) 066017, arXiv:1705.07943.

[306] S. D. Mathur, "How fast can a black hole release its information?," *Int. J. Mod. Phys. D* **18** (2009) 2215, arXiv:0905.4483.